DIPARTIMENTO DI SCIENZE DELL'AMBIENTE FORESTALE E DELLE SUE RISORSE

TESI DI DOTTORATO DI RICERCA IN
ECOLOGIA FORESTALE – XX CICLO

STUDIO SULLA RAPPRESENTATIVITA' DI UNA RETE DI FLUXTOWER CONDOTTO SU SCALA EUROPEA

Settore Scientifico Disciplinare – AGR/05

Dottoranda:
ELEONORA CANFORA

Coordinatore:
Prof. PAOLO DE ANGELIS

Tutori:
Prof. RICCARDO VALENTINI

Dr. DARIO PAPALE

Università degli Studi della Tuscia
Dipartimento di Scienze dell'Ambiente Forestale e delle sue Risorse (DISAFRI)
Via S. Camillo de Lellis, snc 01100 Viterbo

Corso di Dottorato di Ricerca in Ecologia Forestale
Coordinatore: Prof. Paolo De Angelis

Tesi di Dottorato di Ricerca in Ecologia Forestale (XX ciclo)
di: Eleonora Canfora

Gennaio 2009

Indice

1. Introduzione 1

2. Ambito di studio

2.1 Il progetto CarboEuropeIP 6

2.2 I siti di monitoraggio 9

2.3 Il flusso e la standardizzazione dei dati 13

2.3.1 Il metodo di "Spike Detection" 15

2.3.2 "Gap Filling" e "Partitioning" 16

2.4 Il dataset della ricerca 18

3. Analisi mediante tecniche GIS

3.1 Introduzione 20

3.2 Teoria dei GIS: l'acquisizione dei dati 21

3.3 Teoria dei GIS: l'analisi spaziale 22

3.3.1 Classificazioni, aggregazioni, selezioni 23

3.3.2 La sovrapposizione (overlay) 24

3.4 Metodologia applicata 26

3.4.1 Il dominio 26

3.4.2 Il modello di analisi spaziale 27

3.4.3 L'Indice di Coincidenza 28

3.5 Risultati e discussioni 29

3.5.1 Le Homogeneous Territory Unit 29

3.5.2 L'analisi della coincidenza 35

3.5.3 La rappresentatività ecofisiologica 37

3.6 Conclusioni 39

4. Analisi mediante applicazione di SOM

4.1 Introduzione 41

4.2 Teoria di base delle SOM 42

4.2.1 Idea e topologia 42
4.2.2 La regola di apprendimento 45
4.2.3 La visualizzazione 49

4.3 Metodologia applicata 52

4.3.1 Il dominio 52
4.3.2 La struttura SOM applicata 53

4.4 Risultati e discussioni 54

4.4.1 Le mappe SOM 54
4.4.2 La distribuzione geografica 62
4.4.3 I siti teorici 72
4.4.4 Gli "hotspots" 75

4.5 Conclusioni 79

5. Conclusioni finali e sviluppi 82

5.1 Criticità 84

5.2 Prospettive 85

6. Bibliografia 86

Ringraziamenti

Capitolo 1

Introduzione

Dopo anni di sviluppo tecnologico, la società mondiale sta affrontando ora una delle necessità più impellenti e delle sfide più difficili dell'ultimo secolo: stabilizzare i livelli di concentrazione dei gas serra nell'atmosfera per prevenire effetti di riscaldamento sul clima terrestre. In questo ambito, particolare rilievo hanno avuto nell'ultimo decennio gli studi sul ruolo che la vegetazione svolge nel regolare gli scambi di flusso di carbonio tra la superficie terrestre e l'atmosfera. Attraverso un delicato bilancio tra fotosintesi e respirazione, infatti, gli ecosistemi terrestri sono oggi ritenuti in grado di assorbire una parte significativa di anidride carbonica immessa nell'atmosfera (Valentini, 2003). L'accumulo o il rilascio di carbonio è strettamente associato con le diverse tipologie vegetazionali, con i loro mutamenti provocati dalle attività umane, con i cambi di uso del suolo, e con i cambiamenti climatici stessi. Numerosi sono stati gli studi condotti in questo senso e molteplici le tecniche strumentali messe a punto per la misurazione di tali flussi. Tra quelle più utilizzate, per la facilità con cui vengono registrati i dati e per la capacità di effettuare misure in continuo temporale, si è affermata da diversi anni la tecnica cosiddetta Eddy Covariance che si basa sull'analisi della turbolenza atmosferica per la misura degli scambi gassosi (Baldocchi et al., 2001). Dal 1996 ad oggi l'Europa ha visto proliferare numerosi siti di monitoraggio che utilizzano questa tecnica, dislocati in molte delle nazioni europee; essi misurano gli scambi gassosi tra superficie e atmosfera, registrando ormai da un decennio dati su scala semioraria e inviandoli periodicamente ad un Database, sede unica della loro raccolta e della loro standardizzazione. Le torri di monitoraggio sono poste in luoghi estremamente diversificati tra loro, sia per posizione geografica, sia per altitudine, sia per caratteristiche climatiche, sia per tipologia vegetazionale. Fin dalla sua costituzione, che comprendeva solo circa una decina di siti appartenenti al progetto Euroflux, si è posta l'attenzione sull'effettiva capacità della rete stessa di monitorare una situazione che riproducesse in qualche modo il territorio europeo nella sua totalità e complessità. Per questa ragione l'iniziale disposizione dei siti tendeva

comunque a coprire, per quanto ragionevolmente possibile, la distribuzione degli ecosistemi per area e produttività. La nascita di nuovi progetti che ad Euroflux si sono succeduti (Carboeuroflux, Carboeurope), ha dato il via al proliferare di numerosi altri siti che per la maggior parte sono stati in seguito stabiliti in base a necessità scientifiche locali. Questo ha dato vita ad una rete di monitoraggio molto più estesa (circa un centinaio di siti distribuiti sul territorio europeo), la cui capacità di rappresentare le condizioni del territorio stesso, non è stata però ulteriormente verificata.

La rappresentatività non è tuttavia un parametro direttamente misurabile (O'Connor et al., 1990; Awimbo, Norton, 1996) e non è univocamente definibile. Nel suo concetto più generale essa è definita come "…la capacità di un campione di riflettere le caratteristiche di un gruppo di dimensioni maggiori a cui appartiene". Un campione rappresentativo è dunque quello che rispecchia le caratteristiche della popolazione da cui è prelevato e quindi permette di effettuare stime e previsioni attendibili. Calata in ambito ecologico, la rappresentatività di un campione si valuta dalla sua capacità di riprodurre le variabili che caratterizzano un dominio. Quindi non si può disgiungere la rappresentatività dal dominio stesso in cui essa è valutata. Le caratteristiche che definiscono un dominio possono essere di molteplice natura: può esistere un dominio temporale, e in questo caso la rappresentatività potrebbe, per esempio, essere riferita alla capacità del campione di riprodurre nel tempo cambiamenti della popolazione; può esistere un dominio spaziale o geografico, in cui la rappresentatività è valutata in base alla capacità del campione di distribuirsi sul territorio occupato dalla popolazione di provenienza; può esistere un dominio definito da una serie di parametri o di variabili che costituiscono uno spazio a n-dimensioni, in cui il campione si valuta in base alla sua capacità di riprodurre quelle particolari condizioni. Per ogni dominio considerato esiste quindi una rappresentatività diversa, che vale solo in quel caso specifico: un campione può essere rappresentativo di una popolazione per il territorio che occupa, ma non per le caratteristiche fisiche, per alcune variabili la sua rappresentatività può essere notevole e per altri parametri meno, e così via. Esistono quindi rappresentatività innumerevoli, non necessariamente legate ad una posizione geografica o ad una precisa variabile. La misura, o meglio l'analisi della rappresentatività di un campione è quindi strettamente legata al dominio considerato e numerose possono essere le strade scelte per darne una valutazione.

Esempi di studi sulla rappresentatività di aree vegetate, naturali e non, si possono trovare frequentemente applicati nell'ambito della salvaguardia di aree protette (Awimbo et el., 1996; Armenteras D. et al., 2003). In questi studi l'area investigata, il dominio spaziale, è in genere

abbastanza ridotto, su scala locale o al massimo regionale. Su questa scala la valutazione della rappresentatività, anche se pur sempre in funzione delle variabili considerate, risulta in genere più agevole da perseguire. Diventa invece più complessa da definire e da valutare la rappresentatività considerata su scala maggiore, nazionale o continentale. Un importante studio di rappresentatività su un dominio spaziale così esteso è stato condotto negli Stati Uniti sulle torri Eddy Covariance dei siti di monitoraggio afferenti alla rete Ameriflux (Hargrove et al., 2003). Questo studio è stato basato sulla creazione di "ecoregioni" omogenee quanto possibile rispetto alle caratteristiche ambientali incluse nel dominio considerato e al livello di suddivisione stabilito. Le ecoregioni sono state definite usando un *Multivariate Geographic Clustering* e l'analisi condotta ha portato ad una valutazione della rete di monitoraggio in base alla sua capacità di riprodurre le loro caratteristiche, permettendo di individuare all'interno dell'area in esame zone ben rappresentate e zone in cui si renderebbe necessaria l'installazione di nuove torri.

Partendo dall'analisi descritta in questo unico studio di rappresentatività condotto negli Stati Uniti, lo scopo di questo lavoro è quello di dare una prima valutazione sulla rappresentatività della rete dei siti eddy presente sul territorio europeo e afferente al progetto CarboeuropeIP, in termini di caratteristiche climatiche ed ecofisiologiche.
Il dominio rappresentato geograficamente dal territorio europeo presenta tuttavia caratteristiche molto diverse da quelle espresse dal continente nord-americano. L'Europa infatti, pur se di dimensioni relativamente ridotte rispetto agli Stati Uniti, si presenta notevolmente più frammentata a livello di paesaggio e di uso del suolo, e l'individuazione di ecoregioni così come sono state determinate nello studio americano, condurrebbe ad un mosaico troppo variegato per poter essere significativamente analizzato. Si è quindi tentato di valutare la rappresentatività attraverso un approccio diversificato.

Nella prima parte del lavoro viene presentato il progetto nel cui ambito la rete si è sviluppata, le sue finalità, le variabili che ogni torre misura e quelle che costituiscono il dataset di riferimento. Il dominio preso in considerazione è costituito da una componente spaziale (il territorio europeo) e una componente "parametrica", in cui sono state prese in considerazione variabili climatiche legate alla geografia del territorio, e variabili ecofisiologiche legate alla tipologia vegetazionale esistente su di esso. Per descrivere quantitativamente questo dominio

su cui poi vengono confrontati i dati delle torri, sono stati utilizzati dati spazializzati provenienti da diverse fonti: modelli climatici, ecofisiologici, e dati satellitari.
Successivamente ai dati vengono applicate le classiche tecniche di analisi spaziale che sono proprie dei sistemi GIS (Geographical Information System) per individuare in un dominio costituito da un limitato numero di variabili, unità omogeneee di territorio, simili come concetto alle ecoregioni; su queste aree viene valutata la distribuzione geografica dei siti della rete e la loro "coincidenza" con le caratteristiche espresse dalle variabili climatiche ed ecofisiologiche prese in esame. Viene infine calcolata la percentuale di area che nel territorio europeo è occupata da unità di territorio rappresentate dai siti.
Volendo poi ampliare il dominio in cui valutare la rappresentatività , si conduce in seguito un'analisi, descritta nella parte successiva, che si avvale dell'uso di reti neurali auto-organizzanti (SOM, Self Organizing Map). Queste rappresentano un ormai consolidato strumento per la visualizzazione di dati altamente multidimensionali, proiettando relazioni statistiche non lineari complesse in una griglia bidimensionale (Luyssaert, 2007). In questa fase quindi viene condotto uno studio su un dominio costituito da uno spazio a 21 dimensioni, in cui accanto alle variabili climatiche ed ecofisiologiche già considerate nell'analisi GIS, vengono incluse le loro variabilità annuali e stagionali. Inoltre, per poter rapportare in modo più rigoroso i dati puntuali con i dati spazializzati, l'analisi viene condotta separatamente per cinque tipologie vegetazionali per ognuna delle quali è nota, dettagliata e quantificata la distribuzione sul territorio; questa suddivisione è dunque da considerarsi un'ulteriore dimensione del dominio. Le tipologie forestali considerate sono cinque, di cui tre sono di tipo forestale (latifoglie sempreverdi, latifoglie decidue e conifere), mentre le altre sono rappresentate dalle praterie e dai coltivi. Le SOM vengono usate per la creazione di cluster all'interno di ogni singola tipologia che vengono quindi confrontati con le torri della rete di monitoraggio situate al loro interno. Viene data una valutazione comparativa sulla rappresentatività delle singole tipologie vegetazionali e attraverso l'analisi dei cluster vengono individuate le aree in cui ipotizzare una nuova distribuzione di siti.
Seguono le conclusioni finali emerse dallo studio e la bibliografia utilizzata.

Questa ricerca rappresenta quindi un primo approccio nella valutazione della rete di monitoraggio dei siti eddy facenti parte del progetto CarboeuropeIP, dove viene considerata una rappresentatività geografica ed ecofisiologica, che si pone come obiettivo quello di dare

un contributo per migliorare la rete valutandone il disegno e analizzandone la possibilità di miglioramento e di crescita.

Capitolo 2

Ambito di studio

Lo studio è stato condotto nell'ambito del progetto CarboeuropeIP e prende in esame una rete di siti di monitoraggio situati all'interno del territorio europeo.

2.1 Il progetto CarboeuropeIP

Il progetto CarboEuropeIP è supportato dalla Comunità Europea, Direttorato Generale della Ricerca, e rientra nel Sesto Programma Quadro, Priorità 1.1.6.3 Cambiamenti Globali ed Ecosistemi.

Il progetto Carboeurope, da cui si sviluppa il successivo progetto integrato (IP), nasce nel 2000 dall'unione di un gruppo di progetti europei (Medeflux, Euroflux, Carboeuroflux, Tacos, Camels), che nel corso degli anni hanno riunito e consolidato una comunità scientifica interdisciplinare nel campo del monitoraggio di diversi ecosistemi e dell'atmosfera. Il Progetto CarboEuropeIP iniziato nel 2004 riunisce e implementa questi progetti precedenti permettendo per la prima volta una consistente acquisizione di dati armonizzata su tutto il territorio europeo e l'integrazione di diverse scale spaziali e temporali.

Il consorzio attualmente coinvolge 61 Centri di Ricerca situati in 17 diverse nazioni, circa 30 istituti associati all'interno del territorio europeo e si avvale della collaborazione di alcuni istituti extraeuropei. L'Istituto coordinatore del progetto è il Max Planck Institute per la Biogeochimica in Germania.

Lo scopo del progetto è di conoscere e quantificare il bilancio del Carbonio terrestre dell'Europa e di individuare le incertezze del suo monitoraggio su scala locale, regionale e continentale. Per raggiungere questo obiettivo strategico il progetto si focalizza sulle seguenti linee:

- determinare il bilancio del Carbonio europeo con i suoi modelli spaziali e temporali; in questo ambito la chiave innovativa del progetto sta nella sua concezione di applicare singole strategie sperimentali e la loro integrazione in un comprensivo quadro di dati di assimilazione del carbonio. L'aumento della risoluzione spaziale e temporale dei programmi di osservazione e modellizzazione permette per la prima volta

- l'applicazione di approcci sia bottom-up che top-down per determinare il bilancio europeo del carbonio e la variabilità di source e sink presenti negli ecosistemi;
- conoscere i processi di controllo e i meccanismi del ciclo del Carbonio negli ecosistemi europei e come questi sono influenzati dai cambiamenti climatici, dalla variabilità e dalla gestione antropica; questo è raggiunto attraverso la suddivisione dei flussi di carbonio misurati nelle loro parti costituenti (assimilazione, respirazione, emissione di combustibili fossili), e attraverso la quantificazione degli effetti della gestione degli ecosistemi sullo scambio netto di carbonio e attraverso lo sviluppo, la valutazione e l'ottimizzazione di modelli che simulano e prevedono i processi di scambio;
- sviluppare un sistema di osservazione per monitorare i cambiamenti di accumulo di carbonio e dei suoi flussi in relazione agli impegni assunti dall'Europa nell'ambito del protocollo di Kyoto, con lo sviluppo di una proposta di "conteggio" del carbonio basato su misure dei flussi tra superficie e atmosfera, inventari sui cambiamenti di accumulo di carbonio nel suolo e nella biomassa, proprietà della vegetazione monitorata da tecniche di telerilevamento e concentrazioni atmosferiche.

Per raggiungere gli scopi che il progetto si pone, CarboEuropeIP è stato diviso in 4 Componenti:

- Misure a livello ecosistemico;
- Misure atmosferiche a scala continentale;
- Esperimenti a scala regionale finalizzati a ridurre le incertezze nello "scaling";
- Integrazione delle componenti a scala continentale che integra le varie fonti dei dati in un comprensiva valutazione del bilancio del carbonio europeo.

Tutte le componenti interagiscono strettamente tra loro e richiedono informazioni trasversali che le correlano (fig. 2.1). Lo scopo ultimo e comune è quello di stimare il bilancio del carbonio europeo nel recente passato e nel presente.

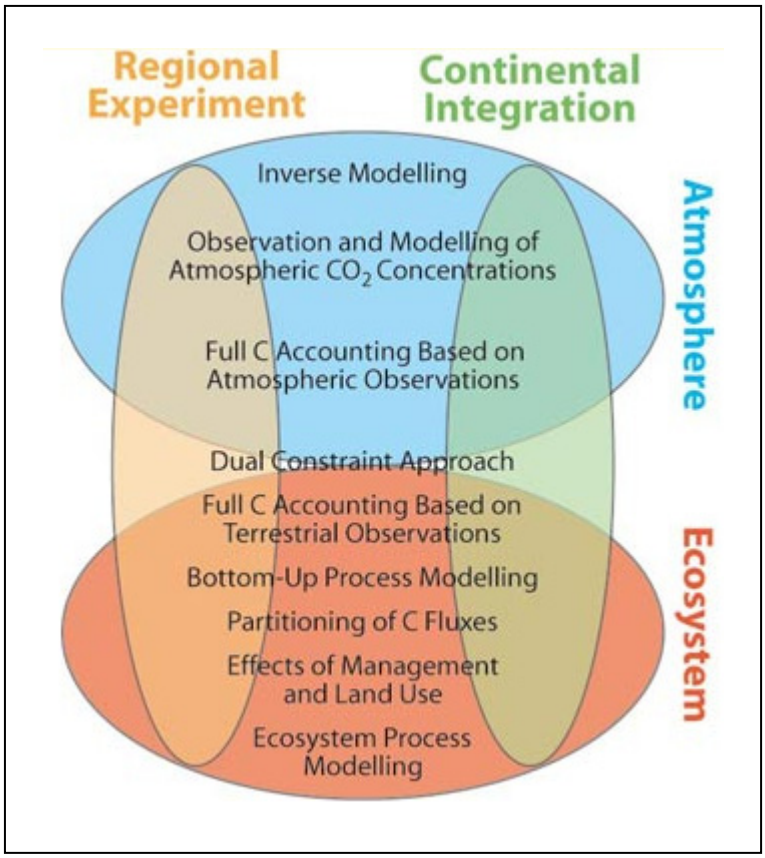


Figura 2.1. Rappresentazione grafica degli obiettivi del progetto Carboeurope e della loro implementazione come componenti del Progetto Integrato (da www.carboeurope.org)

L'attività presentata in questo studio si colloca all'interno della prima delle 4 componenti del progetto, la "Ecosystem Component": essa fornisce la misura dei flussi e il bilancio del carbonio nei principali elementi del paesaggio europeo, riflettendo la loro diversità e la loro variabilità annuale legata al clima e ad altri fattori ambientali, i flussi di carbonio tra pool di un ecosistema (biomassa e suolo), il loro scambio con l'atmosfera, monitorando i processi biogeochimici per una valutazione del bilancio del carbonio europeo in una scala bottom-up.
Gli obiettivi di questa componente del progetto sono: osservare i flussi di carbonio, di acqua e di energia dei principali tipi di coperture vegetali presenti in Europa, mediante la tecnica di Eddy Covariance (Aubinet et al, 2000; Baldocchi et al., 2001); la suddivisione del flusso netto di carbonio nelle sue parti costituenti in base tipo di uso del suolo e alla sua gestione, usando scale temporali che variano da quelle orarie a quelle annuali o pluriannuali; quantificare gli effetti della gestione dei siti di monitoraggio sul flusso netto di carbonio e lo scambio di CH_4 e N_2O; fornire i dati che sono alla base di valutazioni e parametrizzazioni di modelli di up-scaling di flussi di carbonio a scala regionale e continentali. Ed è appunto in quest'ultimo obiettivo della componente del progetto, che si inserisce il lavoro presentato.

Una particolare attività di ricerca è volta a risolvere anche problemi di tipo metodologico, per assicurare controlli standardizzati sulla qualità dei dati e per la calibrazione tra i diversi siti della rete di monitoraggio (par 2.3).

2.2 I siti di monitoraggio

Nel progetto CarboeuropeIP la rete dei siti di monitoraggio è costituita di 51 Siti Principali (Main Sites, che includono 12 siti intensivi che forniscono anche misure di accumulo di carbonio nel suolo e turnover), e circa 50 Siti Associati (Associated Sites) (fig. 2.2).

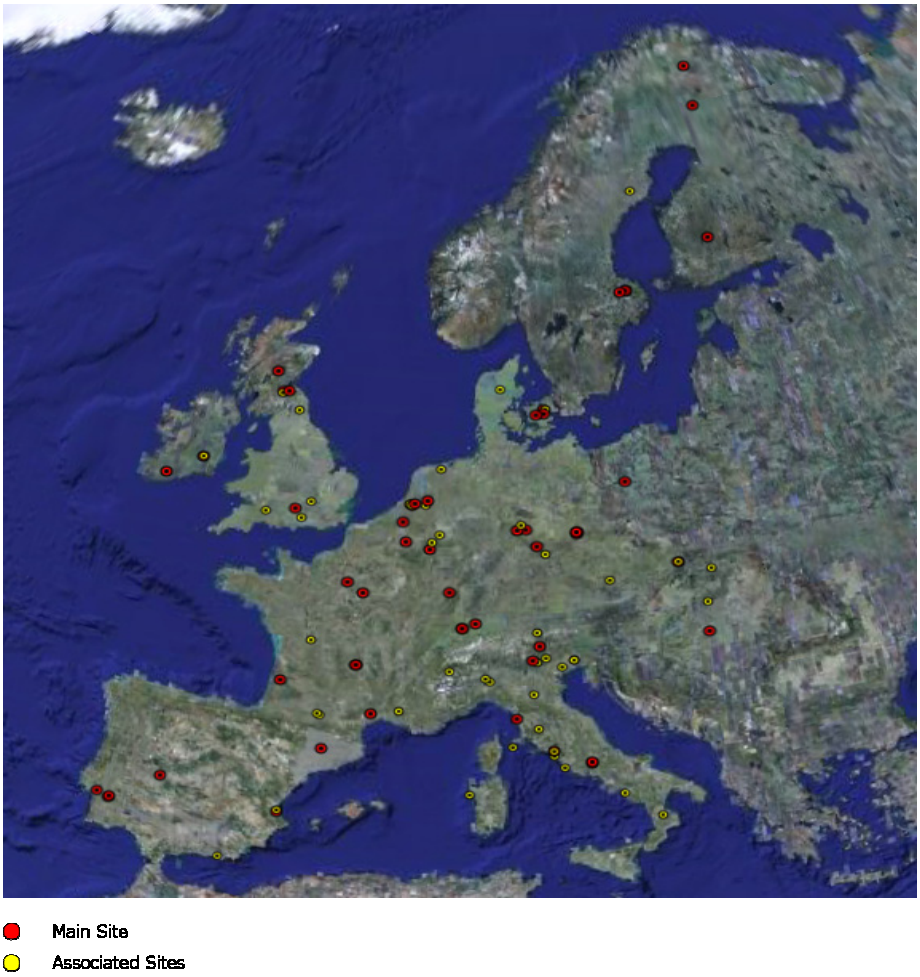


Figura 2.2. La rete dei siti di monitoraggio di CarboeuropeIP

La distribuzione dei siti di studio dovrebbe coprire ragionevolmente la distribuzione degli ecosistemi per area e per Net Biome Productivity (NBP, Janssens et al., 2001). La rete è composta da 16 cluster che dovrebbero idealmente anche coprire i principali tipi di uso del

suolo di un paesaggio. Un cluster è un insieme di siti di almeno tre principali tipi di ecosistemi, che consiste di tre o più Siti Principali e altri Siti Associati. I Main Site devono fornire al progetto dati continui per almeno 5 anni, mentre gli Associated Site possono inviare dati anche solo per un anno. Nelle tabelle 2.1a e 2.1b sono elencati tutti i siti appartenenti al progetto; per ognuno di essi viene indicato il codice con cui è identificato all'interno del progetto stesso, il nome del sito, il Plant Functional Type e, dove è possibile, il tipo di copertura del suolo. In figura 2.3 è mostrata la rete dei siti distinti in base al loro PFT.

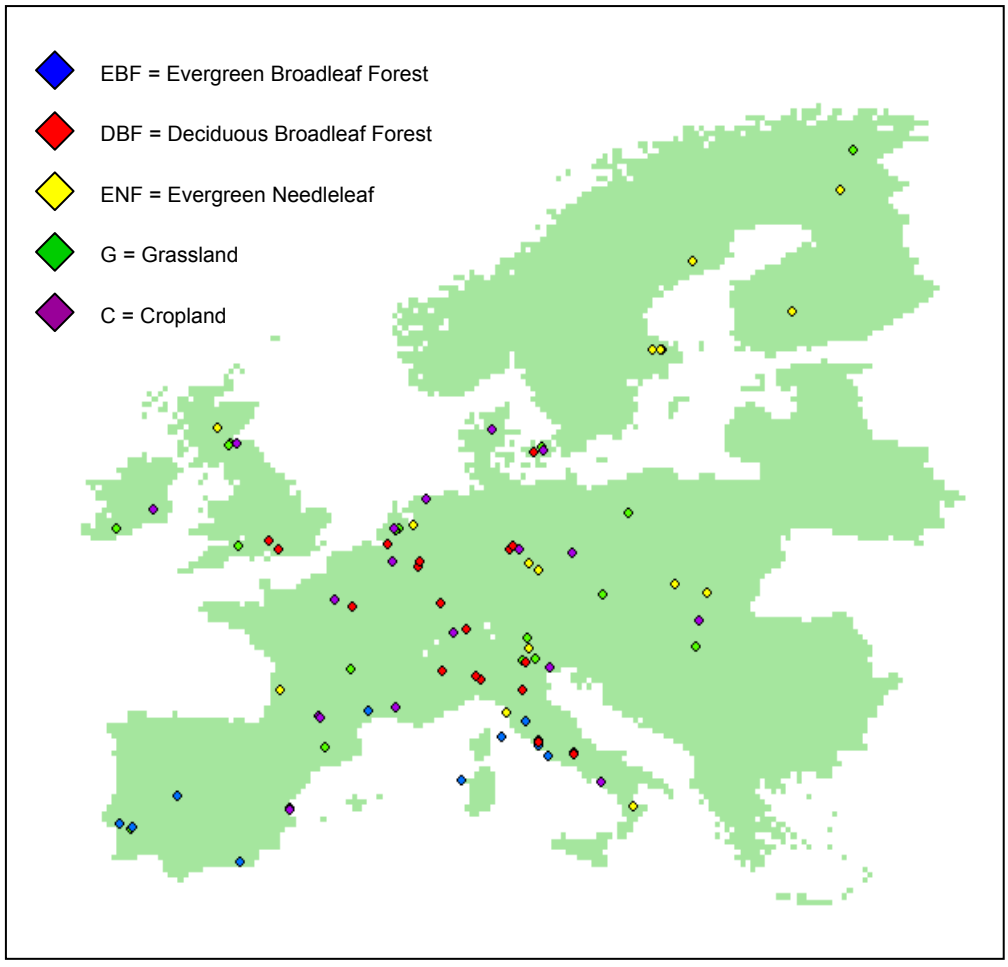


Figura 2.3. Distribuzione geografica dei siti per PFT

CEIP Code	Site Name	PFT	Land Cover
BE-Bra	Brasschaat	DBF	Pinus Sylvestris+Quercus robur
BE-Lon	Lonzee	C	Arable crop (rotation)
BE-Vie	Vielsalm	DBF	Fagus sylvatica, Pseudotsuga
CH-Lae	Laegern	DBF	Mixed forest
CH-Oe1	Oensingen grassland	G	Grassland
CH-Oe2	Oensingen crop	C	Arable
CZ-BK1	Bily Kriz forest	ENF	Coniferous forest
DE-Geb	Gebesee	C	Cropland
DE-Gri	Grillenburg	G	meadow
DE-Hai	Hainich	DBF	beech
DE-Kli	Klingenberg	C	crop rotation
DE-Tha	Tharandt	ENF	spruce
DE-Wet	Wetzstein	ENF	spruce
DK-Lva	Rimi	G	Grassland
DK-Ris	Risbyholm	C	Cropland
DK-Sor	Soroe	DBF	beech
ES-ES2	El Saler-Sueca (Valencia)	C	Cropland
ES-LMa	Las Majadas del Tietar (Caceres)	EBF	Dehesa (Open Oakland)
ES-VDA	Vall d' Alinyà (Lleida)	G	Grassland
FI-Hyy	Hyytiala	ENF	scots pine 41 yr
FI-Kaa	Kaamanen	W	grassland (peat)
FI-Sod	Sodankyla	ENF	scots pine 100 yr
FR-Fon	Fontainbleu	DBF	oak
FR-Gri	Grignon	C	Cropland
FR-Hes	Hesse	DBF	beech
FR-LBr	Le Bray	ENF	pine
FR-Lq1	Laqueuille intensive	G	Grassland
FR-Lq2	Laqueuille extensive	G	Grassland
FR-Pue	Puechabon	EBF	oak
HU-Bug	Bugac	G	Grassland
IE-Ca1	Carlow crop	C	Arable
IE-Dri	Dripsey	G	Grassland
IT-Amp	Amplero	G	Grassland
IT-Col	Collelongo	DBF	beech
IT-MBo	Monte Bondone	G	Grassland
IT-Ren	Renon	ENF	spruce
IT-Ro1	Roccarespampani 1	DBF	oak
IT-SRo	San Rossore	ENF	pine
NL-Ca1	Cabauw	G	Grassland
NL-Hor	Horstermeer	G	natural fen (peat)
NL-Loo	Loobos	ENF	Scots Pine
PL-wet	POLWET	W	wetlands
PT-Esp	Espirra	EBF	Eucalyptus plantation
PT-Mi1	Mitra II (Evora)	EBF	oak woodlands
PT-Mi2	Mitra IV (Tojal)	G	Grassland
SE-Nor	Norunda	ENF	scots pine/spruce
SE-Sk1	Skyttorp 1	ENF	scots pine (0 y)
SE-Sk2	Skyttorp 2	ENF	scots pine (30 y)
UK-EBu	Easter Bush	G	Grassland
UK-ESa	East Saltoun	C	barley/grass
UK-Gri	Griffin	ENF	Sitka Spruce
UK-PL3	Pang/Lambourne forest	DBF	forest broadleaf

Tabella 2.1a. Main Site del progetto CarboeuropeIP (PFT = Plant Functional Type)

CEIP Code	Site Name	PFT	Land Cover
AT-Neu	Neustift	G	Grassland
BE-Jal	Jalhay	DBF	mixed forest, planted
CZ-BK2	Bily Kriz grassland	G	Grassland
CZ-wet	CZECHWET	W	wetlands
DE-Meh	Mehrstedt	DBF	afforestation
DE-Seh	Selhausen	C	Cropland
DK-Eng	Enghave	G	Grassland
DK-Fou	Foulum	C	Cropland
ES-ES1	El Saler (Valencia)	ENF	High maquia + pine
ES-LJu	Llano de los Juanes	EBF	Olive Plantation
FR-Aur	Auradé	C	Crop rotation
FR-Avi	Avignon	C	crop
FR-Lam	Lamasquère	C	Crop rotation
FR-Lus	Lusignan	G	grassland
HU-Mat	Matra	C	Cropland
IE-Ca2	Carlow grassland	G	Carlow grassland
IT-BCi	Borgo Cioffi	C	Cropland
IT-Be1	Beano1	G	Cropland
IT-Be2	Beano2	C	Grassland
IT-Bon	Bonis	ENF	F/plantation
IT-Cas	Castellaro	C	Cropland
IT-Cpz	Castelporziano	EBF	oak
IT-Ctv	Castelvetrano	C	Cropland
IT-Lav	Lavarone	DBF	spruce
IT-Lec	Lecceto	EBF	oak
IT-LMa	La Mandria	DBF	oak
IT-Mal	Malga Arpaco	G	Grassland
IT-Neg	Negrisia - Ponte di Piave	C	Cropland
IT-Noe	Arca di Noé - Le Prigionette	EBF	Mediterranean Macchia
IT-Non	Nonantola	DBF	mixed dec
IT-Pia	Pianosa	EBF	shrub/crop
IT-PT1	Parco Ticino Forest	DBF	poplar
IT-Ro2	Roccarespampani 2	DBF	oak
IT-Tol	Tolfa	EBF	shrub
IT-Vig	Vigevano	C	Short Rotation Forestry
NL-Ca2	Cabauw extension	G	arable land/grassland
NL-Dij	Dijkgraaf	C	Cropland
NL-Haa	Haastrecht	G	Grassland
NL-Lan	Langerak	C	Maize
NL-Lel	Lelystad	G	Grassland
NL-Lut	Lutjewad	C	Winter wheat
NL-Mol	Molenweg	C	Potatoes
NL-Zee	Zeewolde	C	Cropland
SE-Fla	Flakaliden	ENF	Norway spruce, 35 years
SK-Tat	Tatra Danielov Dom	ENF	Coniferous Forest
UK-AMo	Auchencorth Moss	G	Blanket peat
UK-Ham	Hampshire	DBF	Forest (broadleaf)
UK-Har	Harwood	ENF	Sitka Spruce
UK-Her	Hertfordshire	C	Arable
UK-PL1	Pang/Lambourne crop	C	Arable
UK-PL2	Pang/Lambourne grassland	G	Grassland
UK-Tad	Tadham Moor	G	forest broadleaf

Tabella 2.1b. Associated Site del progetto CarboeuropeIP (PFT = Plant Functional Type)

2.3 Il flusso e la standardizzazione dei dati

Tutti i dati raccolti dai siti di monitoraggio vengono inviati ad un Database della Ecosystem Component che provvede alla loro standardizzazione, resa necessaria per ottenere un dataset omogeneo e al suo interno comparabile. Ogni sito eddy raccoglie dati di flusso e dati meteorologici (tab. 2.2 a e b)

Flux Variables

Variable	Description	Unit
CO2	carbon dioxide concentration	μmol/mol
H2O	water vapour concentration	mmol/mol
ZL	atmosphere stability parameter	Dimensionless
Fc	carbon dioxide	μmol m -2 s-1
qcFc	quality class	Dimensionless
H	sensible heat	W m-2
qcH	quality class	Dimensionless
LE	latent heat	W m-2
qcLE	quality class	Dimensionless
τ	momentum	Kg m-1 s-2
qcTau	quality class	Dimensionless
Ustar	friction velocity	m s-1
TR	tree transpiration	ml hr -1
Sb	canopy heat storage	W m-2
Sc	CO2 storage in canopy air layer	μ mol m-2 s-1
Sw	latent heat in canopy air layer	W m-2
Sa	heat storage in canopy air layer	W m-2

Tabella 2.2a. Variabili di flusso misurate nei siti eddy

Meteorological Variables

Variable	Description	Unit
P	precipitation	Mm
Rg	global or short wave incoming radiation	W m-2
Rr	reflected or short wave outgoing radiation	W m-2
LWin	long wave incoming radiation	W m-2
LWout	long wave outgoing radiation	W m-2
Rn	net radiation	W m-2
Rd	diffuse radiation (global)	W m-2
PPFD	photosynthetic photon flux density	μ mol m-2 s-1
PPFDd	diffuse photosynthetic photon flux density	μ mol m-2 s-1
PPFDr	reflected photosynthetic photon flux density	μ mol m-2 s-1
PPFDbc	below-canopy photosynthetic photon flux density	μ mol m-2 s-1
APAR	light interception	μ mol m-2 s-1
Ta	air temperature	°C
Pa	pressure	Kpa
Tc	canopy radiative temperature	°C
Tbole	bole temperature	°C
Ts1	soil temperature superficial	°C
Ts2	soil temperature medium	°C
Ts3	soil temperature deep	°C
SWC1	soil water content superficial	%vol
SWC2	soil water content medium	%vol
SWC3	soil water content deep	%vol
G1	soil heat flux	W m-2
G2	soil heat flux	W m-2
Rh	relative humidity	%
WD	Wind direction	degrees
WS	Wind horizontal speed	m sec-1

Tabella 2.2b. Variabili meteorologiche misurate nei siti eddy

Il dato registrato a questo punto necessita di una serie di controlli e quality check volti ad eliminare le misure affette da errori strumentali o registrate in condizioni meteorologiche non idonee all'applicazione della tecnica della correlazione turbolenta. La metodologia utilizzata per testare la qualità dei dati provenienti dalle stazioni di misura è quella prevista a livello internazionale secondo Foken (Foken et al., 2003).Una volta che i dati sono stati importati, vengono processati per migliorare la qualità del dato stesso usando tecniche standardizzate di Quality check descritte in Papale et al. (2006). Durante il Quality Check vengono effettuati due tipi di test sui dati di flusso (Fc): il primo per identificare i valori di bassa variabilità (*low*

variability test) e il secondo per individuare i punti marginali, cioè tutti quei punti che delimitano gap di dati, che possono essere affetti da errore.

Dopo aver ottimizzato l'analisi della qualità dei dati per inserire valori di "*quality flag*", ai dati stessi viene applicato un modello di "spike detection", passaggio necessario per preparare le serie di dati alle tecniche di "*gapfilling*".

2.3.1 Il metodo di "spike detection"

Le misure effettuate con la tecnica Eddy Covariance sono spesso influenzate dagli spike dovuti a fattori biofisici e strumentali. La presenza degli spike è relazionata a diversi aspetti e in particolare alle caratteristiche del sito e dall'opera di "ripulitura" operata dal responsabile della torre.
Gli spike relativi alla misura puntuale vengono eliminati in fase di acquisizione prima di calcolare il valore relativo alla mezz'ora, mentre rimangono nel dataset gli spike nella serie annuale dei valori (la media della mezz'ora è uno spike rispetto alle medie della mezz'ora precedente e successiva). Questi spike non influenzano direttamente il valore di NEE che viene calcolato dai dati, ma possono influenzare la qualità dei dati gapfilled prodotti in un passaggio successivo. Il metodo usato per individuare gli spike si basa sulla posizione del valore di ciascuna mezz'ora rispetto a quello della mezz'ora precedente e successiva ed è applicato ad una finestra di 13 giorni, separatamente per valori diurni e notturni.

Per ogni valori di NEE calcolato in una mezz'ora si definisce un parametro *d* tale che:

$$d_i = \left(NEE_i - NEE_{i-1}\right) - \left(NEE_{i+1} - NEE_i\right) \qquad \text{eq. 2.1}$$

e il valore è individuato come spike se:

$$d_i < Md - \left(\frac{z \cdot MAD}{0.6745}\right) \qquad \text{eq. 2.2}$$

oppure se

$$d_i > Md + \left(\frac{z \cdot MAD}{0.6745}\right) \qquad \text{eq. 2.3}$$

dove *Md* è la media delle differenze, mentre *MAD* è definita come:

$$MAD = media\left(\left|d_i - Md\right|\right) \qquad \text{eq. 2.4}$$

e *z* è un valore soglia.

Per valutare i dati vengono assegnati a *z* diversi valori, in dipendenza anche della sensibilità del metodo che si vuole ottenere. Nel caso dei dati inseriti nel database le soglie stabilite sono 3: $z = 4$, che è il valore usato convenzionalmente, $z = 5.5$ e $z = 7$ che sono valori che chiaramente danno un despiking più conservativo.

2.3.2 "Gap Filling" e "Partitioning"

Una volta applicato il metodo di despiking, i dati vengono sottoposti a tecniche di gap filling e di partitioning, l'applicazione dei quali permette di ottenere dai valori di NEE l'Assimilazione (Gross Ecosystem Production, GEP) e la Respirazione (Ecosystem Respiration, R_{eco}), dell'ecosistema secondo l'equazione

$$NEE = R_{eco} - GEP \qquad \text{eq. 2.5}$$

La "partizione" di NEE è necessaria per approfondire in quale misura i suoi valori dipendano dalla variabilità annuale all'interno di ogni sito e tra siti diversi. La conoscenza di questa dipendenza di NEE da fattori diversi è importante per l'applicazione delle tecniche di gap filling che vengono applicate ai dati ogni volta che le misure vengono registrate sotto condizioni meteorologiche sfavorevoli o durante il malfunzionamento degli strumenti, permettendo così il calcolo del budget annuale di NEE.

L'algoritmo (MDS) applicato all'intero dataset del progetto CarboeuropeIP, è descritto in Reichstein, (2005). Esso considera una sensibilità a breve termine che la respirazione dell'ecosistema presenta in funzione della temperatura e la applica per estrapolare valori di R_{eco} diurni da quelli notturni; per ciascun punto la R_{eco} viene calcolata secondo :

$$Reco(t) = (t)\, e^{(E0(1/(Tref-T0) - (1/(Tsoil(t)-T0))} \qquad \text{eq. 2.6}$$

L'algoritmo stesso inoltre sviluppa un processo di gap filling che utilizza sia la covarianza tra i dati di flusso e meteorologici, sia la struttura temporale dei flussi stessi. Questi ultimi vengono inoltre corretti e filtrati utilizzando lo "*storage*" e la soglia di ustar che viene calcolata secondo un criterio simile a quello usato in Reichstein et al., (2002). In questo algoritmo tre sono le situazioni prese in considerazione:

1) risultano mancanti alcuni dati di interesse, ma sono presenti tutti i dati meteorologici

2) non sono disponibili anche la temperatura o il VPD ma sono presenti valori di radiazione
3) non sono disponibili anche i valori di radiazione

Nel primo caso i valori vengono sostituiti con la media calcolata in una finestra di $\pm$ 7 giorni nelle stesse condizioni meteorologiche. Lo stesso procedimento viene applicato anche nel secondo caso ma nelle condizioni di simile Rg (con una deviazione al di sotto di 50 W m^{-2}). Nel caso in cui non siano disponibili i valori di Rg, il gap viene riempito con valore medio calcolato allo stesso giorno ($\pm$ 1h), in questo caso la finestra considerata è $\pm$ 0.5 giorni. Se alla fine del procedimento i valori non possono comunque essere calcolati, si ripete l'algoritmo aumentando la dimensione delle finestre fino a quando risulta possibile riempire il *gap* dei dati mancanti. In figura 2.4 viene mostrato il diagramma di flusso dell'intero algoritmo.

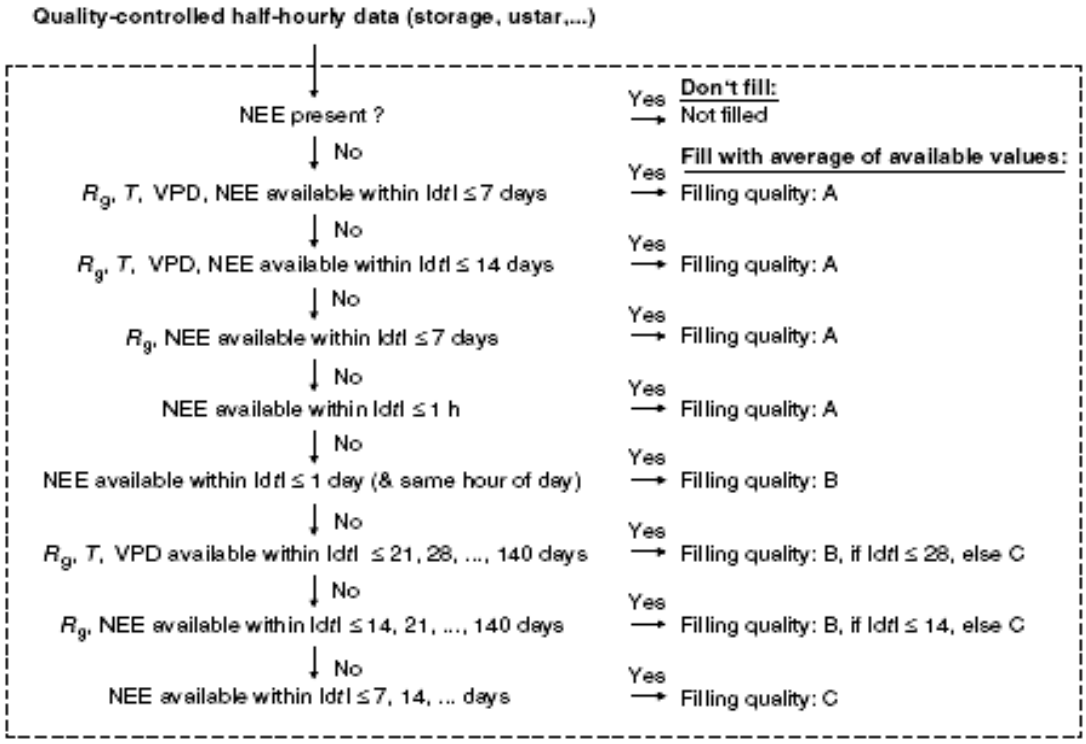


Figura 2.4. Diagramma di flusso dell'algoritmo applicato per il calcolo del *Gap Filling*

Il prodotto risultante dall'applicazione dell'algoritmo è rappresentato da una serie di 4 file dove la NEE, la GPP e la R_{eco} vengono calcolate su scala semi oraria, giornaliera, settimanale e mensile.

Il gap filling infine viene condotto utilizzando anche le reti neurali (ANN) secondo Papale et al. (2006) e il confronto tra i risultati ottenuti dai due metodi (MDS e ANN) fornisce una ulteriore base per l'analisi della qualità dei dati.
Le correzioni e le standardizzazioni effettuate sui vari dataset permettono a questo punto di valutare la qualità dei dati di NEE calcolati a partire da misure effettuate con la tecnica "eddy covariance", passaggio necessario per un' analisi comparativa tra i dati provenienti da diversi siti e per l'uso dei dati stessi in studi su scala maggiore come quello presentato in questo lavoro.

2.4 Il dataset della ricerca

Il dataset è quello che determina il dominio in cui la ricerca è stata condotta: insieme alla componente spaziale (il territorio europeo), viene considerata la componente climatica e la componente eco fisiologica.
Accanto ai dati che provengono dai siti appartenenti al progetto CarboeuropeIP, standardizzati così come descritto nel paragrafo 2.3, vengono raccolti dati spazializzati che sono in grado di descrivere il dominio europeo in cui valutare la rappresentatività della rete di monitoraggio.
Le variabili ecofisiologiche, NEE e GPP sono state desunte dal modello di spazializzazione che usa le reti neurali per determinare gli scambi di carbonio su scala europea, modello descritto in Papale e Valentini (2003). I dati su scala europea sono forniti distinti per Plant Functional Type.
Le variabili meteorologiche sono invece ottenute dal dataset REMO (Regional Model) fornito dal Max Planck Institute per la Meteorologia. Da qui sono stati estratti dati spazializzati su scala europea di Radiazione Globale, Temperatura media, Precipitazione, Vapour Deficit Pressure.
Vengono infine aggiunti dati di frazione fotosinteticamente attiva (fpar) ottenuti dai dati satellitari del MODIS (Moderate Resolution Imaging Spectroradiometer), di proprietà della NASA. Anche i dati di fpar sono distinti per Plant Functional Type.
Tutti i dati spazializzati coprono l'intero territorio europeo con una immagine avente le seguenti caratteristiche:

Angolo superiore sinistro: Long. 15° W, Lat. 75° N
Angolo inferiore destro: Long. 60° E, Lat. 30°N

Risoluzione spaziale: Griglia di 300 righe e 180 colonne (risoluzione della cella 0.25 x 0.25 gradi)

Per poter infine analizzare le singole tipologie vegetazionali, al dataset viene aggiunta una mappa di Land Cover fornita dal sistema MODIS che permette di assegnare per ogni cella del territorio europeo, la percentuale di copertura appartenente ad uno specifico PFT.

In tabella 2.3 viene riassunto il dataset usato per l'analisi:

	Variabile	Fonte	Risoluzione temporale
Dati dai siti di monitoraggio (periodo 2000 - 2006)	GPP	CarboeuropeIP	giornaliera
	NEE	CarboeuropeIP	giornaliera
	Rg	CarboeuropeIP	giornaliera
	Ta	CarboeuropeIP	giornaliera
	VPD	CarboeuropeIP	giornaliera
	Prec.	CarboeuropeIP	giornaliera
Dati Spazializzati (periodo 2000 - 2006)	GPP	modello ANN*	giornaliera
	NEE	modello ANN	giornaliera
	Rg	REMO**	giornaliera
	Ta	REMO	giornaliera
	VPD	REMO	giornaliera
	Prec.	REMO	giornaliera
	fpar	MODIS***	8 giorni
	Land Cover	MODIS***	

* Papale, Valentini (2003)
** Regional Model - Max Planck Institute per la Meteorologia
*** Moderate Resolution Imaging Spectroradiometer - NASA

Tabella 2.3. Dataset utilizzato nello studio

Capitolo 3

Analisi mediante tecniche GIS

3.1 Introduzione

Per poter valutare la rappresentatività dei dati puntuali derivanti dai singoli siti della rete di monitoraggio (capitolo 1) nel dominio costituito dall'insieme delle variabili considerate e descritte nel capitolo 2, che sono spazializzate nel territorio europeo preso in esame, è necessario fare ricorso a tecniche in grado di gestire dati riferiti allo spazio geografico. In questa parte della ricerca, ai dati della rete di monitoraggio e al dominio a cui si riferiscono, vengono applicate le classiche tecniche che si avvalgono dell'uso dei sistemi GIS (Geographical Information System).

Un GIS è un sistema che contiene specifiche funzioni per acquisire, elaborare, gestire e restituire dati georeferiti. I sistemi GIS nascono e si sviluppano dalla seconda metà degli anni '60 e acquisiscono potenzialità e importanza man mano che aumentano la produzione sistematica e la disponibilità di cartografia digitale, la costituzione di banche dati geografiche e il potenziamento di software. Oggi i sistemi GIS sono applicati in modo ormai consolidato in molteplici campi, dall'economico al gestionale, a quello scientifico.

I moderni GIS sono in grado di gestire dati di cartografia numerica georeferenziata (in coordinate metriche), dati descrittivi direttamente associati agli elementi rappresentati sulla cartografia, ma anche qualsiasi altro dato indirettamente georeferibile. Essi si distinguono da altri sistemi di progettazione, cartografia e gestione dei database, per la capacità di integrare i dati georeferiti attraverso strumenti di analisi quali: la selezione e la ricerca spaziale, la sovrapposizione automatica di livelli geografici (map overlay), la generazione di aree di rispetto o di influenza (buffer), la generazione di modelli digitali del terreno. La loro finalità è quindi volta all'analisi integrata di informazioni georeferite al territorio, al fine di mostrare i dati acquisiti in vari modi e secondo vari punti di vista (Biallo, 2005).

Per poter svolgere il ruolo ad essi assegnato, essi devono comprendere una serie di funzionalità (Knapp, 1978) che prevedono: l'acquisizione dei dati, che viene svolta in funzione dei risultati che si vogliono ottenere, in cui ogni tipologia di informazione deve

essere classificata, deve essere individuata la fonte e la qualità necessaria, la struttura informatica e la procedura di acquisizione; la pre-elaborazione dei dati, che permette di manipolare i dati in modo da predisporli per il loro inserimento nella banca dati (conversioni fra strutture, georeferenziazioni, conversioni fra proiezioni e coordinate, interpolazioni, foto interpretazioni); l'analisi spaziale che include operatori che permettono di generare e derivare nuove informazioni dei dati contenuti nel sistema; la generazione di prodotti, che comprende la produzione di report statistici e cartografie di vario tipo.

Scopo di questo capitolo della ricerca è quello di applicare tecniche di analisi spaziale per classificare il dominio determinato dal territorio e dai parametri considerati, in modo da suddividerlo in classi omogenee su cui proiettare i dati provenienti dalla rete di monitoraggio. Il confronto tra i dati delle torri e i dati spazializzati viene quindi usato per valutare come i dati che caratterizzano un sito coincidano o meno con quelli presenti sull'area da esso occupata. Da questa informazione si deduce quali aree del dominio hanno una coincidenza maggiore con i dati che provengono dalle torri eddy. Si vuole infine verificare se la rappresentatività geografica coincide o meno con quella determinata esclusivamente dai parametri ecofisiologici considerati ("rappresentatività ecofisiologica").

3.2 Teoria dei GIS: l'acquisizione dei dati

L'oggetto principale che viene rappresentato e studiato all'interno di un GIS è il territorio nella sua complessità, identificato come un sistema reale composto da sottoinsiemi, oggetti e relazioni (Biallo, 2005). L'analisi delle informazioni da inserire nel GIS permette di progettare un modello del mondo reale attraverso l'inserimento di dati opportunamente strutturati e relazionati e di sviluppare procedure di analisi spaziale che simulano fenomeni ed eventi.

Il dato può essere acquisito in diversi modi: rilievo sul campo, foto aeree, digitalizzazione cartografica, telerilevamento da satellite. Per poter analizzare insieme dati derivanti da differenti fonti è però necessario che siano nella stessa proiezione, nello stesso sistema di coordinate ed anche di una qualità compatibile (risoluzione).

Le proiezioni geografiche sono necessarie per poter riprodurre le coordinate geografiche di Latitudine e Longitudine della superficie ellissoidica della Terra in un piano bidimensionale. Non è però possibile rappresentare su un piano una superficie curva senza introdurre delle deformazioni di distanza, di area o di angoli e in questa impossibilità i sistemi di

rappresentazione tendono ad ottimizzare i tre parametri, privilegiandone uno. Si parla così di proiezioni equidistanti (mantengono inalterate le distanze), equivalenti (risultano inalterate le aree), e conformi (mantengono inalterati gli angoli). Esistono in questi tre ambiti numerosi tipi specifici di proiezioni: prospettiche, coniche, cilindriche ecc.

La costruzione di una carta prevede però, oltre all'applicazione di una proiezione geografica, anche l'utilizzo di un sistema di riferimento che permette di individuare una corrispondenza tra le coordinate geografiche (Lat e Long) e le coordinate del piano (x, y) in modo tale che ad ogni punto dell'ellissoide corrispondano due coordinate metriche, cosa che consente di calcolare distanze, angoli, aree. Un sistema di riferimento è definito infine dal *datum*, che individua il modello matematico usato per calcolare le coordinate geografiche dei singoli punti. Numerosi sono i sistemi di riferimento adottati e spesso la scelta dipende dall'area che si vuole investigare: esistono infatti sistemi di riferimento "locali" che vengono applicati quando l'analisi coinvolge piccole aree a scala nazionale (i sistemi di riferimento più utilizzati in Italia sono per esempio il Roma 40-Gauss-Boaga, che comprende il territorio nazionale in due fusi, Est e Ovest, di 6° di ampiezza, e l'European Datum 50 che comprende il territorio nazionale in due fusi di 6° di ampiezza, denominati fuso 32 e 33 del sistema internazionale) e altri che vengono maggiormente usati per studi a scala continentale o mondiale. Fra i sistemi di riferimento internazionali, introdotti dalle tecnologie satellitari, viene molto spesso usato il WGS84 (World Geodetic System, 1984); il centro dell'ellissoide coincide con il centro della massa della Terra, l'asse Z è diretto verso il Polo Nord, l'asse X è diretto verso il meridiano zero di Greenwhich e l'asse Y completa una terna cartesiana destrorsa. A questo sistema di riferimento non è ufficialmente associato alcun sistema di proiezione cartografica.

3.3 Teoria dei GIS: l'analisi spaziale

La funzione di analisi spaziale è la caratteristica principale che identifica un sistema GIS da altre tipologie di Sistemi Informativi in grado di operare su dati geografici.

Le funzioni di analisi spaziale permettono di elaborare dati geografici e descrittivi per rispondere a domande specifiche sull'oggetto della ricerca. La significatività dell'analisi spaziale è ancora più evidente quando si abbia la necessità di operare su scala continentale o quando il modello di analisi è così complesso da rendere necessaria l'integrazione di più dati e l'uso di diverse funzioni, oppure quando si abbia la necessità di analizzare vari scenari e diverse alternative, oppure infine quando l'operazione sia ripetitiva su dati continuamente

aggiornati. Perché i risultati siano attendibili, l'analisi spaziale richiede una attenzione particolare sull'organizzazione dei dati e sulla sequenza delle analisi ed i risultati; la qualità dell'informazione prodotta dipende dall'approccio sistematico dell'analisi e quindi dall'uso dei dati e degli strumenti (DeMers, 2005).

In generale comunque l'analisi spaziale permette di estrarre informazioni da dati esistenti che siano per forma e contenuto utili allo scopo, oppure di generare nuovi oggetti geografici partendo da informazioni già in possesso, oppure ancora di analizzare modelli particolari di oggetti geografici quali quelli tridimensionali o di rete. Secondo le esigenze di analisi si può utilizzare una sola funzione o la combinazione di più funzioni.

3.3.1 Classificazioni, aggregazioni, selezioni

Le funzioni di riclassificazione, aggregazione e selezione possono operare solo su dati descrittivi oppure su dati geografici e descrittivi insieme.

Un esempio di *riclassificazione* consiste nella generazione di un nuovo attributo descrittivo partendo da un attributo già esistente: partendo dalla fonte originaria dei dati si può ipotizzare di dividerli in classi e di attribuire a ciascuna classe un nuovo codice da inserire in un nuovo attributo; questo può avere la finalità di semplificare la visualizzazione delle classi oppure di facilitare processi successivi di elaborazioni ed analisi. E' possibile anche generare un attributo dall'interazione di più attributi preesistenti; attraverso i nuovi attributi generati è possibile visualizzare facilmente i dati riclassificati, eventualmente utilizzarli per ulteriori elaborazioni, oppure generare nuovi file contenenti solo gli elementi geografici di una certa classe.

Le classificazioni possono essere di tipo supervised e unsupervised: nel primo caso la classificazione viene condotta a partire da gruppi di oggetti che vengono precedentemente riconosciuti sul campo, nel secondo caso la suddivisione in classi viene stabilita in base ad analisi statistiche multivariate che vengono applicate ai dati senza che se conosca a priori la loro distribuzione.

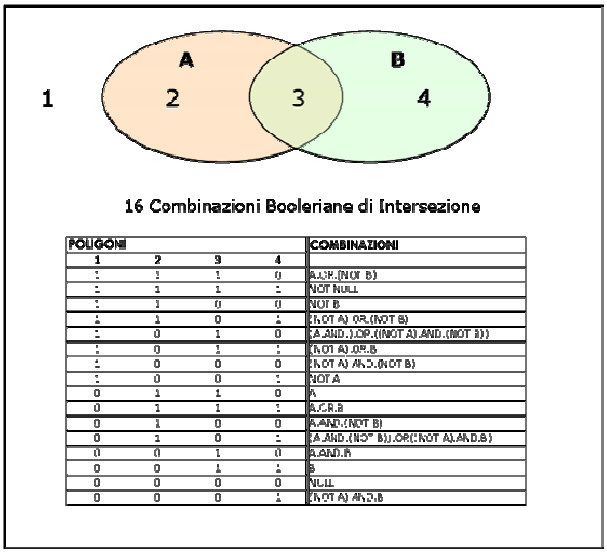
Una classe corrisponde a un significativo gruppo di posizioni, ciascuna posizione è caratterizzata da un set di valori per una specifica variabile. Ciascuna posizione quindi può essere visualizzata come un punto sullo spazio multidimensionale i cui assi corrispondono alle variabili rappresentate. Due posizioni occupano la stessa classe se i loro valori sono simili.

Una volta classificati i dati è possibile anche operare con ulteriori funzioni, quali per esempio l'*aggregazione* fisica dei poligoni che appartengono a due classi distinte, eliminando, dove si presentano, le linee di separazione. Una volta generato il nuovo attributo che classifica con lo stesso codice entrambe le classi, si passa ad operare sul dato geografico, applicando una funzione topologica di aggregazione che elimina (dissolve) i limiti di separazione fra i poligoni della stessa classe generata, riducendone il numero e generando nuovi classi omogenee. La funzione di aggregazione è tipica dei dati vettoriali (dati che sono rappresentati fisicamente come punti, linee e poligoni, o un insieme composto di queste entità grafiche di base), mentre per i dati raster (dove il territorio è suddiviso secondo un reticolato ortogonale regolare in celle di forma quadrata ad ampiezza uniforme, numerate secondo gli elementi di una matrice) è sufficiente un'operazione di riclassificazione per ottenere un nuovo file dove scompaia la suddivisione delle classi di partenza.
Un altro esempio di funzione che viene spesso applicata dopo una classificazione è la *selezione spaziale* che permette di evidenziare o estrarre elementi geografici sulla base di poligoni definiti interattivamente o selezionati da altri file. La selezione può comportare anche la generazione di un nuovo file che contenga solo gli oggetti geografici che soddisfano determinate condizioni.

3.3.2 La sovrapposizione (overlay)

Le funzioni di sovrapposizione di oggetti geografici (dette "overlay") permettono di generare una nuova mappa cartografica che sintetizza elementi geografici ed attributi ricavati dai dati di partenza. Le operazioni di overlay possono essere eseguite sia su dati raster che su dati vettoriali, sovrapponendo livelli informativi differenti; nel caso si sovrappongano due livelli raster, la funzione di overlay genera un nuovo raster i cui valori di pixel sono ricavati dai valori dei corrispondenti pixel, di posizione omologa, dei raster di partenza. Naturalmente l'operazione può essere eseguita anche su più raster contemporaneamente; il valore di ciascun pixel del nuovo raster generato può essere il risultato di una qualsiasi operazione eseguita sui pixel omologhi dei raster di partenza: la somma, la media, il valore minimo, il massimo, un valore ricavato da una matrice di classificazione.
Le funzioni di overlay sui dati vettoriali permettono invece di generare nuovi informazioni vettoriali dall'intersezione di differenti livelli di oggetti geografici.

L'overlay di una classe di oggetti poligonali ed una classe di oggetti lineari, per esempio, darà come risultato un nuovo file di oggetti lineari, frazionati nei punti di intersezione fra i poligoni e le linee; i tratti di linee ricadenti in ciascun poligono avranno, in un nuovo attributo, il codice del poligono stesso. L'overlay fra livelli poligonali costituisce la funzione che, insieme al buffer, caratterizza inequivocabilmente uno strumento GIS come tale; la sovrapposizione fra livelli poligonali differenti può dare sedici diversi risultati secondo l'espressione booleriana adoperata (fig.3.1).



POLIGONI				COMBINAZIONI
1	2	3	4	
1	1	1	0	A.OR.(NOT B)
1	1	1	1	NOT NULL
1	1	0	0	NOT B
1	1	0	1	(NOT A) OR.(NOT B)
1	0	1	0	(A.AND.).OR.((NOT A).AND.(NOT B))
1	0	1	1	(NOT A) .OR.B
1	0	0	0	(NOT A) AND.(NOT B)
1	0	0	1	NOT A
0	1	1	0	A
0	1	1	1	A.OR.B
0	1	0	0	A.AND.(NOT B)
0	1	0	1	(A.AND.(NOT B)).OR.(NOT A).AND.B)
0	0	1	0	B.AND.A
0	0	1	1	B
0	0	0	0	NULL
0	0	0	1	(NOT A) AND.B

Figura 3.1. Possibili combinazioni derivanti da un'operazione di overlay topologico

Molto usate sono A.OR.B (generalmente detta "unione"), e A.AND.B (generalmente detta "intersezione"). Il nuovo livello geografico, grazie agli identificatori degli oggetti di origine presenti sui due livelli differenti, può ricevere in eredità tutti gli attributi degli oggetti di origine che si vuole; inoltre le aree dei poligoni così generati saranno riferite alle nuove geometrie assunte.

Uno degli usi delle funzioni di overlay è quello di "tagliare" un file di oggetti geografici sulla base di un'area regolare o irregolare definita interattivamente o estratta da un altro file; operazione analoga è quella di tagliare un file di poligoni secondo un rettangolo o un area definita dall'utente come area di studio.

L'overlay vettoriale però nasconde delle insidie da non sottovalutare; infatti sovrapponendo aree si possono generare micro poligoni con aree infinitamente piccole e nodi di intersezione a volte invisibili; poiché un poligono per quanto piccolo, genera un record nella tabella

topologica degli attributi, si può risalire ad essi attraverso l'area; anche le linee così generate risulteranno molto piccole ed i nodi intersezione molto vicini. In genere quindi si opera con funzioni di editing utilizzando tolleranze e selezioni in base alle dimensioni. Molta attenzione va posta nella scelta delle operazioni di ripulitura automatica, in quanto non è possibile controllare tutti i punti e tutti i casi in cui vengono effettuate dal software modifiche geometriche, che possono inficiare l'uso e la qualità del dato elaborato.

3.4 Metodologia applicata

3.4.1 Il dominio

Il dominio "parametrico", associato a quello spaziale relativo al territorio europeo, in cui valutare la rete dei siti di monitoraggio considerato in questa parte della ricerca, è costituito da 5 variabili scelte tra quelle ecofisiologiche e climatiche (tabella 2.3) che generalmente vengono misurate anche a livello delle singole torri eddy.

GPP	gC m^{-2} y^{-1}
Precipitazioni annue (Precip.)	mm y^{-1}
Radiazione globale (Rg)	MJ m^{-2} y^{-1}
Temperatura media (Ta)	°C
Vapour Pressure Deficit (VPD)	KPa

I dati spazializzati per ognuna delle variabili acquisiti dall'applicazione del modello ANN (Papale, Valentini, 2003) o dal REMO Database, sono stati riferiti al sistema di riferimento WGS84, in modo tale che ad ogni pixel dell'immagine avente le caratteristiche già descritte nel paragrafo 2.4, corrisponda una coppia di coordinate geografiche.
I dati giornalieri sono stati mediati sull'anno e per ognuna delle variabili sono state calcolate le medie interannuali secondo le equazioni:

Media interannuale: $m = \bar{x}\left(\sum annuale\right)$ per GPP e Precip eq. 3.1

$m = \bar{x}\left(\bar{x}\ annuale\right)$ per Rg, Ta, VPD eq. 3.2

Per l'acquisizione dei dati provenienti dalle torri dei siti di monitoraggio, sono stati considerati solo quei siti per i quali sono pervenuti al CarboeuropeIP Database dati per almeno un intero anno solare, per un totale di 89 siti. Da ogni sito, se presenti più anni di dati, sono stati esclusi dall'analisi eventuali dati provenienti da anni non completi.
Anche per i dati puntuali delle torri sono state calcolate le medie annuali e quindi quelle interannuali secondo le equazioni 3.1 e 3.2.

3.4.2 Il modello di analisi spaziale

I dati spazializzati (formato raster) vengono classificati mediante un algoritmo di classificazione unsupervised che per la definizione delle classi si basa sull'algoritmo Natural Breaks di Jenks (Jenks, Caspall, 1971). Questo schema di classificazione determina la migliore suddivisione in classi comparando le somme delle differenze dei quadrati tra i valori osservati entro ciascuna classe e le medie delle classi stesse (eq. 3.3). La migliore classificazione è quella che identifica le divisioni nella distribuzione dei valori in modo tale da minimizzare la somme delle differenze dei quadrati entro ciascuna classe.

$$SSD_{i\ldots j} = \sum_{k=i}^{j} \left(A[\ k\ ] - mean_{i\ldots j} \right)^2 \qquad \text{eq. 3.3}$$

Secondo questo algoritmo quindi, le divisioni tra le classi sono determinate statisticamente trovando coppie adiacenti di gruppi tra cui c'è però una relativamente ampia differenza di valori. Le classi cioè sono basate sui gruppi naturali contenuti nei dati. Vengono identificati i limiti tra una classe e l'altra selezionando i range che meglio raggruppano valori simili massimizzando però la differenza tra le classi. I gruppi sono quindi divisi in classi i cui bordi sono posizionati dove c'è appunto una elevata differenza nei valori. Per ogni variabile considerata, la classificazione restituisce una mappa dove ad ogni pixel è attribuito un numero relativo alla classe a cui appartiene.
Una volta classificata ognuna delle 5 variabili prese in esame, viene eseguita un'operazione di overlay tra le mappe risultanti, in cui i valori attribuiti a ciascuna classe vengono uniti tra loro in tutte le loro possibili combinazioni con una funzione appunto detta di "Combine". Questa funzione, che appartiene alla classe "not null" delle combinazioni Booleriane di Intersezione (dove quindi non viene escluso nessun valore dall'analisi di intersezione dei poligoni),

assegna un unico valore a ciascuna combinazione di valori trovata e lo attribuisce ad ogni pixel di una mappa (fig. 3.2).

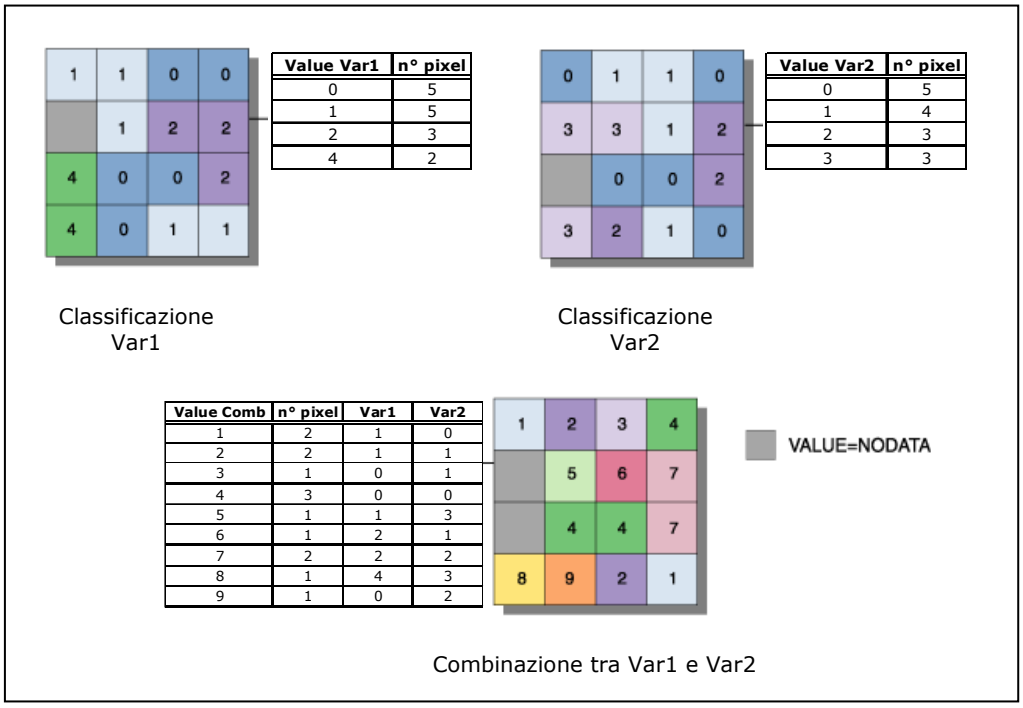


Figura 3.2. Esempio di funzione "Combine"

Sui dati puntuali delle stazioni (formato vettoriale) viene quindi effettuata la stessa analisi eseguita per i dati spazializzati. Viene loro imposta la stessa classificazione, individuando per ogni variabile gli stessi range individuati dall'algoritmo Natural Breaks Jenks sui dati spazializzati, e vengono create le combinazioni tra le varie classi.

Campionando la mappa delle combinazioni dei dati spazializzati con le coordinate dei punti si ottengono così due dataset (le combinazioni estratte dai dati spazializzati e quelle calcolate sui dati puntuali delle torri) completamente paragonabili l'uno con l'altro su cui poi viene condotto il confronto.

3.4.3 L'Indice di Coincidenza

Una volta assegnate le combinazioni ai pixel estratti dai dati spazializzati e alle medie delle variabili misurate dai siti di monitoraggio, viene condotto il confronto per valutare quanto il sito coincide con l'area che lo contiene. La misura di questa coincidenza è data da un indice

calcolato in base alle differenze tra le classi che derivano dalla classificazione del sito e da quello dell'area in cui cade, espresse per ogni variabile.

Dato un singolo pixel *i* in cui cade un sito *m*, si avrà per questo pixel una combinazione di 5 classi, ognuna relativa ad una variabile:

$A = [\ a_i\ \ b_i\ \ c_i\ \ d_i\ \ e_i\ ]$

e per il sito una analoga combinazione:

$B = [\ a_m\ \ b_m\ \ c_m\ \ d_m\ \ e_m\ ]$

Viene calcolata per ogni variabile la differenza tra il valore del pixel e quello del sito:

$C = [\ \alpha_{im}\ \ \beta_{im}\ \ \gamma_{im}\ \ \delta_{im}\ \ \varepsilon_{im}\ ]$ dove per esempio $\alpha_{im} = |a_i - a_m|$

Considerando che ogni variabile ha lo stesso peso nella valutazione della coincidenza e che la GPP è divisa in 3 classi, la precipitazione in 2 classi, la Rg in 3 classi, la Ta in 3 classi ed infine la VPD in 2 classi, si definisce l'Indice di Coincidenza come:

$$IC = [1-(1/15*\ \alpha_{im})-(1/10*\ \beta_{im})-(1/15*\ \gamma_{im})-(1/15*\ \delta_{im})-(1/10*\ \varepsilon_{im})] \quad \text{eq. 3.4}$$

L'indice ha valori compresi tra 0 e 1 dove 1 indica il 100% di coincidenza delle due combinazioni di variabili considerate.

3.5 Risultati e discussioni

3.5.1 Le Homogeneous Territory Unit

Ognuna delle 5 variabili che costituiscono il dominio parametrico considerato all'interno del territorio europeo viene classificata come descritto nel paragrafo 3.4.2: il risultato è una serie di 5 mappe riportate nella figura 3.3. La classificazione delle variabili prevede l'individuazione di 2 o 3 classi per ciascuna variabile: in particolare vengono create 3 classi

per la GPP, 2 classi per le precipitazioni, 3 classi per la Radiazione Globale, 3 classi per la Temperatura e 2 classi per il VPD.

L'esiguo numero delle classi per ogni variabile è determinato dalla risoluzione dei dati di partenza a disposizione dell'analisi e dalla estrema variabilità del territorio europeo: i pixel di ogni immagine hanno una risoluzione di 0.25 x 0.25 gradi e considerando una classificazione su scala continentale, un numero maggiore di classi avrebbe portato ad un mosaico estremamente frammentato, sul quale sarebbe stato impossibile individuare aree omogenee di significative dimensioni, impedendo così di fatto l'analisi geografica del territorio. Anche con le classi così individuate si possono tuttavia descrivere le principali variazioni nel territorio europeo per ciascuna variabile. Nella classificazione di GPP si evidenzia la grande zona continentale europea che risulta più omogenea rispetto alla zona mediterranea e a quella della penisola scandinava, in quella delle precipitazioni si evidenziano la zona della penisola scandinava quella irlandese e quella dell'arco alpino, la classificazione di Ta distingue non solo la penisola scandinava e l'arco alpino, ma anche la zona mediterranea. Rg e VPD risultano chiaramente più uniformi nella loro distribuzione.

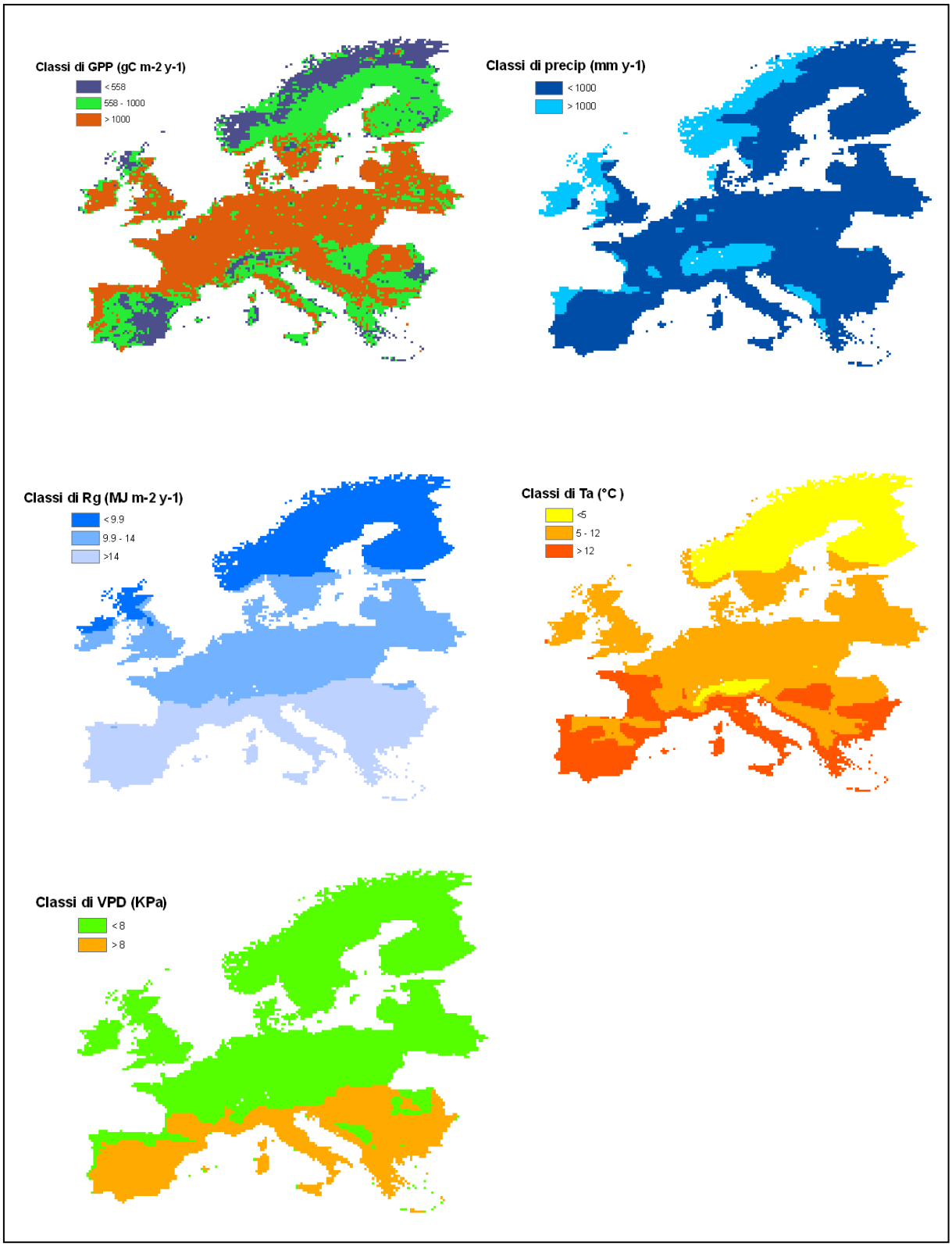


Figura 3.3. Classificazioni unsupervised per GPP, Precip., Rg, Ta, VPD

Una volta classificate le 5 variabili, viene eseguita la loro combinazione secondo la funzione "combine" descritta nel paragrafo 3.4.2. Il risultato è la mappa riportata in figura 3.4:

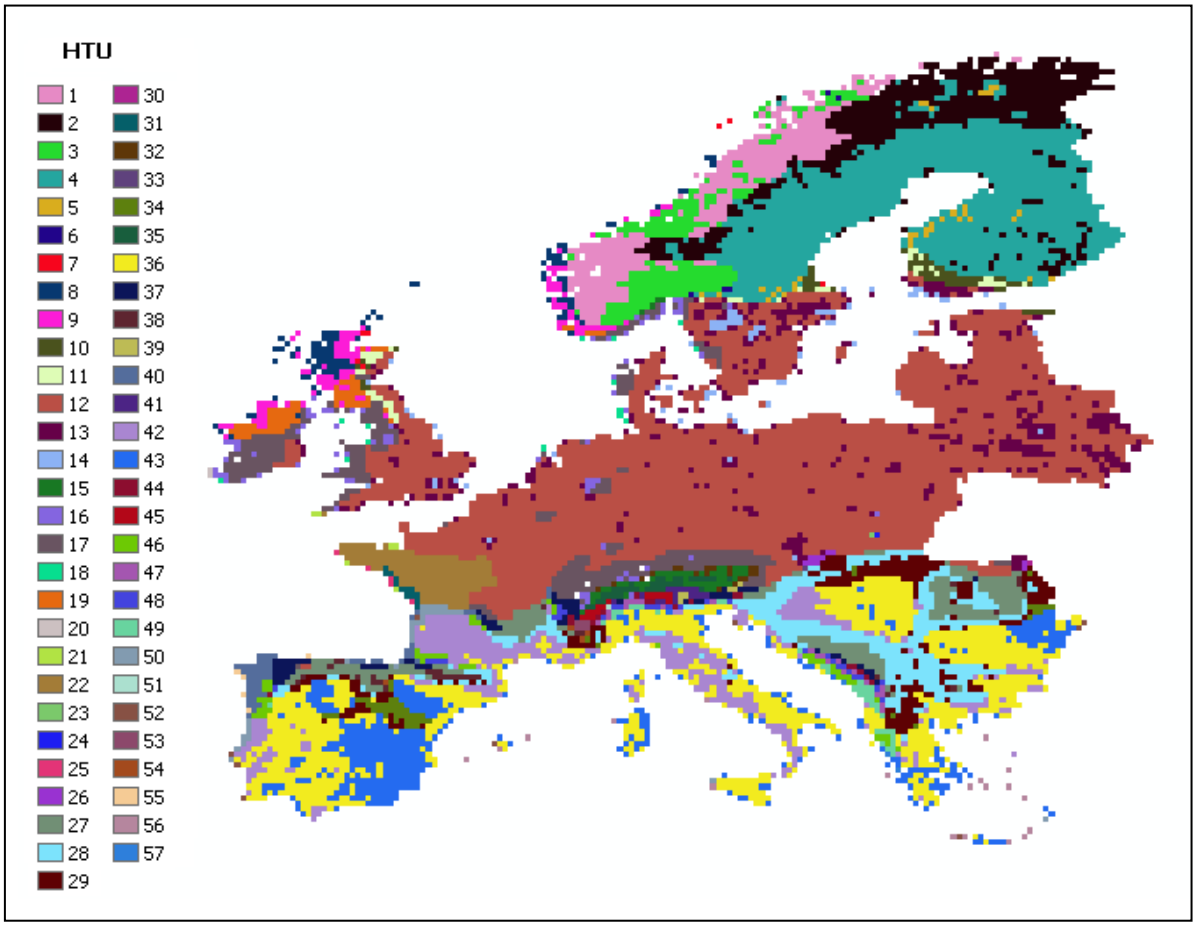


Figura 3.4. Distribuzione geografica delle Homogeneous Territory Unit

Il territorio europeo risulta così suddiviso in 57 aree, le Homogeneous Territory Unit, ognuna delle quali è contraddistinta da una specifica combinazione delle 5 variabili prese in esame. Si evidenzia una estesa HTU che occupa la zona dell'Europa centrale (HTU 12), si distinguono le HTU che caratterizzano la regione scandinava (HTU 1, 2 e 4), quelle che caratterizzano l'est dei Balcani (HTU 27, 28 e 29) e si individuano le HTU che compongono la zona mediterranea (HTU 36, 42, 43) che chiaramente risulta essere in generale la regione più frammentata dell'intero territorio, insieme con l'arco alpino che comprende sul suo territorio più di 10 piccole HTU. Le HTU in cui viene divisa la maggior parte dell'area mediterranea sono determinate dalla classificazione della GPP che in queste aree risulta avere valori meno omogenei, grazie ad un paesaggio maggiormente diversificato e alla coesistenza di tipologie

vegetazionali molto diverse tra loro. Nella tabella 3.1 vengono elencate le HTU create a partire dalle classificazioni e la loro estensione, calcolata secondo Taramelli et al. (in review).

HTU	n° pixel	Area (Km2)	HTU	n° pixel	Area (Km2)
1	415	137649	30	1	518
2	602	177394	31	13	6887
3	264	93365	32	20	10481
4	1320	442516	33	1	521
5	36	12321	34	55	31456
6	5	1696	35	35	18543
7	7	2610	36	1078	619802
8	98	39152	37	76	41909
9	109	44828	38	32	17973
10	66	25207	39	6	3192
11	45	17996	40	61	34397
12	3201	1506170	41	16	8526
13	438	195467	42	458	258820
14	73	30678	43	489	288011
15	53	27370	44	18	9870
16	78	35847	45	15	8063
17	387	187294	46	52	29515
18	14	6179	47	23	12771
19	77	33934	48	31	17136
20	2	952	49	39	22238
21	6	3062	50	32	17949
22	181	94579	51	1	538
23	1	506	52	9	5226
24	1	506	53	6	3279
25	5	2624	54	4	2221
26	10	5168	55	5	2852
27	270	147693	56	20	12150
28	420	231125	57	3	1776
29	254	139405			
			Area Totale		5129909

Tabella 3.1. Estensione delle HTU

Delle 57 HTU solo 14 coprono un'estensione maggiore di 50000 Km^2, a conferma del fatto che pur utilizzando classificazioni non estremamente dettagliate, il territorio europeo si dimostra estremamente eterogeneo.

Proiettando sulla mappa delle HTU i siti di monitoraggio della rete, si ottiene la distribuzione mostrata in figura 3.5. Delle 57 HTU in cui è stato suddiviso il territorio, 18 contengono almeno un sito di monitoraggio, ma la distribuzione è abbastanza difforme: degli 89 siti totali 40 ricadono nell'area più estesa, quella appartenente alla HTU 12, e 28 siti appartengono invece alle HTU 36, 42 e 43 che caratterizzano la zona mediterranea. I restanti 21 siti sono distribuiti nelle altre 14 HTU. Considerando la sola distribuzione geografica e sommando le aree che contengono al loro interno almeno un sito, si ottiene un valore pari al 74.78% dell'area totale del territorio. Questo dimostra una buona distribuzione della rete all'interno del dominio spaziale considerato.

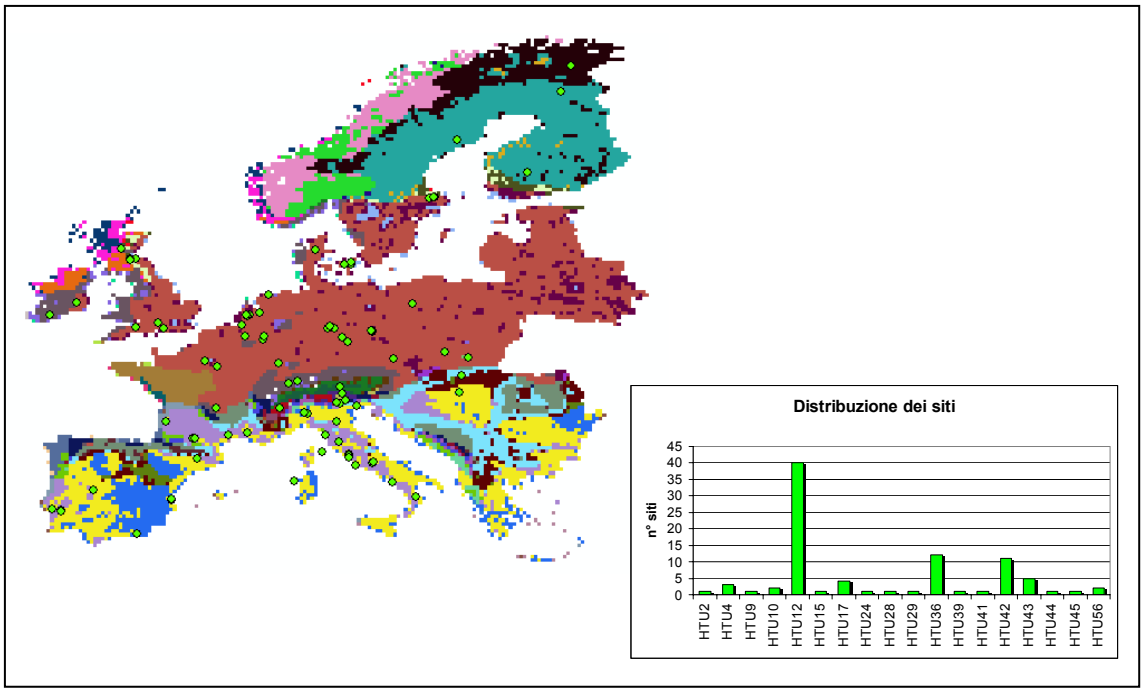


Figura 3.5. Distribuzione dei siti all'interno delle HTU

Confrontando la distribuzione dei siti con l'estensione delle HTU si nota tuttavia come esistano delle estese aree che non vengono coperte da alcun sito (fig. 3.6); in particolare queste zone si collocano in Norvegia, in Irlanda del Nord e a nord della Francia (HTU 1, 3, 19, 22 nella figura 3.4).

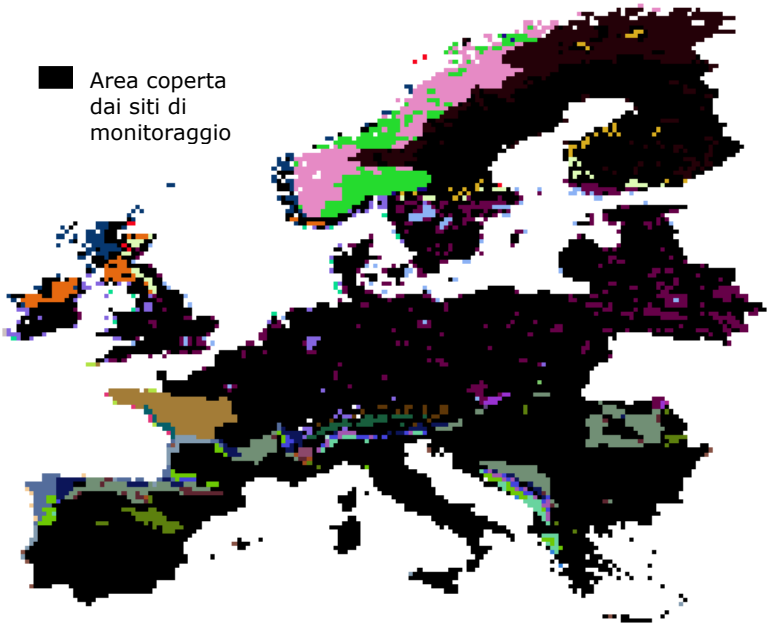


Figura 3.6. HTU prive di siti di monitoraggio

In queste zone potrebbero essere previsti nuovi siti in una ipotetica nuova disposizione della rete di monitoraggio. Altre HTU prive di siti ma di dimensione più piccole si posizionano dove la distribuzione geografiche delle HTU stesse risulta più frammentata, sull'arco alpino, sul nord della Spagna, e sul nord del Regno Unito.

3.5.2 L'analisi della coincidenza

Calcolando l'Indice di Coincidenza (IC) per ogni sito rispetto alla HTU che lo contiene (eq. 3.1), si ottengono i risultati mostrati nel grafico di figura 3.7:

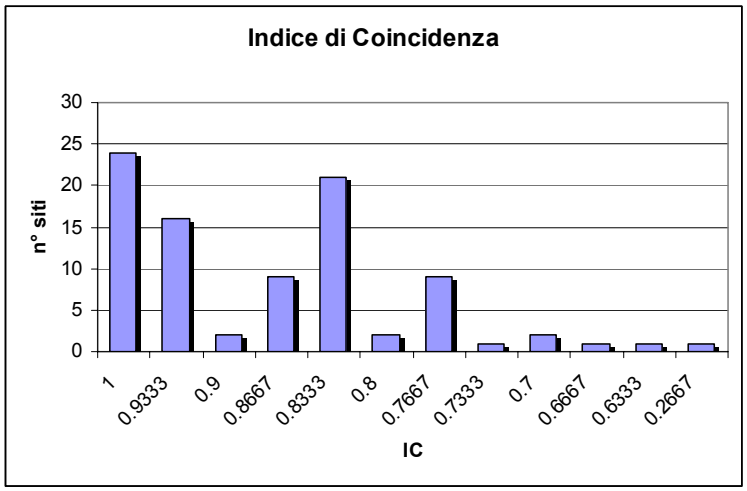


Figura 3.7. Indici di Coincidenza per i siti della rete

degli 89 siti considerati, 42 siti hanno un IC maggiore di 0.9 (47.19% dei siti totali) e 74 (83.15%) hanno un Indice di Coincidenza superiore ad un valore di 0.8 rispetto alle HTU che li contengono, mentre solo il restante 16.85% ha un valore di IC inferiore (solo un sito ha un valore di IC molto basso, pari a 0.267). Questo mostra un buon grado di rappresentatività dei siti sulle aree che li contengono anche secondo il dominio ecofisiologico e climatico considerato.

L'Indice di Coincidenza viene determinato in base alle 5 variabili che nell'equazione espressa nel paragrafo 3.4.3 hanno un peso proporzionato al numero di classi in cui ogni variabile è classificata (eq. 3.4). Analizzando il numero di siti che presentano le diverse differenze tra le classi, rappresentate dai valori di $\alpha, \beta, \gamma, \delta, \varepsilon,$ così come vengono calcolate nello stesso paragrafo 3.4.3 per ogni variabile, si ottengono i risultati mostrati nella tabella 3.2:

Diff.	α (GPP)	β (Prec)	γ (Rg)	δ (Ta)	ε (VPD)
0	52	76	51	75	63
1	29	13	36	12	26
2	8	0	1	1	0
3	0	-	1	1	-

Tabella 3.2. Numero di siti per differenze di classi

Da questa distribuzione emerge che tra le variabili che sono state divise in 3 classi, la GPP risulta quella che influisce più spesso sulla diminuzione di IC dei siti rispetto alle HTU di appartenenza. Sugli 89 siti considerati si notano infatti per questa variabile ben 8 siti che hanno una differenza tra le classi pari a 2 e che nel calcolo dell'Indice di Coincidenza influiscono quindi maggiormente. Anche la Rg ha un numero elevato di siti che vengono classificati in modo diverso dalla HTU (38 siti con γ diverso da 0 contro i 37 della GPP con α diverso da 0), ma la maggior parte di questi presenta una differenza delle classi pari a 1 e solo due siti hanno differenze di classi pari a 2 o 3. In generale quindi la Rg incide in misura minore sul calcolo degli Indici di Coincidenza.
Per le variabili che sono state classificate con l'uso di due classi soltanto (precipitazioni e VPD), la Vapour Pressure Deficit è quella che risulta avere un peso maggiore nell'IC perché per essa ci sono 26 siti che hanno differenze con la classificazione di HTU, mentre per le precipitazioni i siti con queste differenze sono solo 13.

Osservando infine per ogni HTU il numero di siti che mostrano un Indice di Coincidenza maggiore di 0.8 (fig.3.8) si può notare come appunto 13 delle 18 HTU che contengono siti, abbiano almeno un sito con questo IC; in particolare assume rilievo la HTU 12 per la quale tutti i 40 siti contenuti hanno un alto IC. Questo risultato è da mettere chiaramente in relazione con l'elevata omogeneità della HTU che quindi anche con la risoluzione di 0.25 x 0.25 gradi risulta più facilmente rappresentabile. Se invece consideriamo una soglia di IC pari a 0.9, le HTU che contengono siti strettamente correlati risultano essere 8 sulle 18 che contengono siti.

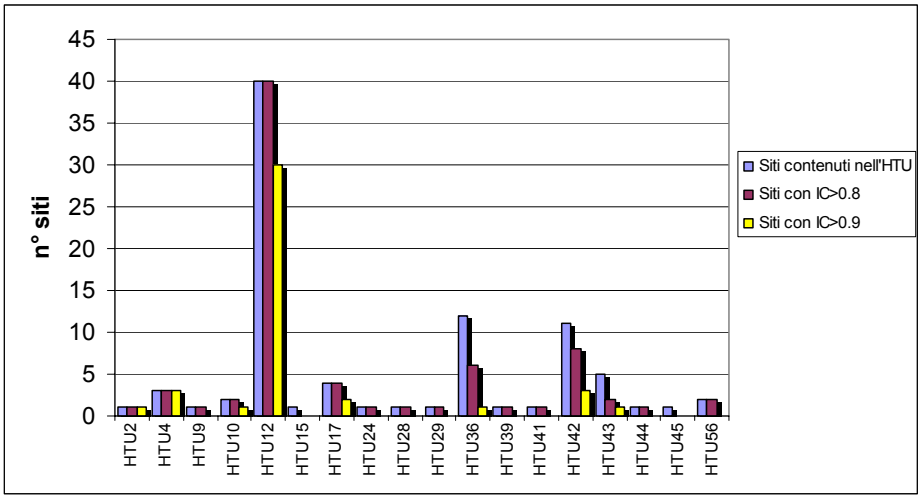


Figura 3.8. Siti contenuti in una HTU con IC>0.8 e con IC>0.9

Solo in 3 HTU non tutti i siti contenuti hanno un IC>0.8: queste tre HTU sono la 36, 42, 43 che sono già state individuate come le HTU che caratterizzano la zona mediterranea (par. 3.5.1). La presenza in questa area di siti con un minore IC può essere messa in relazione con l'alta variabilità che caratterizza questo ambiente, che può aumentare la probabilità che il sito sia posizionato in un punto che non presenta le stesse caratteristiche del pixel che lo contiene. La stessa situazione si presenta anche in due HTU, la 15 e la 45, che non hanno siti con IC maggiore di 0.8: queste HTU sono geograficamente posizionate sull'arco alpino per il quale era già stato messo in evidenza uno stato di alta frammentazione (par. 3.5.1).

In questa analisi appare quindi evidente che la risoluzione dei dati spazializzati utilizzati nello studio (0.25 x 0.25 gradi) può risultare insufficiente là dove i parametri ecofisiologici e climatici considerati nella costituzione del dominio, definiscono aree meno omogenee, mentre risulta più adeguata dove si considerino aree più omogenee.

3.5.3 La rappresentatività ecofisiologica

In ultima analisi sono state confrontate le combinazioni che caratterizzano ciascuna HTU con le combinazioni dei siti della rete di monitoraggio, a prescindere dalla loro posizione geografica, senza cioè considerare se il sito appartiene o meno alla HTU con cui viene confrontato. Per ogni sito è stato calcolato l'Indice di Coincidenza con tutte le HTU generate dalla combinazione delle classificazioni e poi considerati i siti che hanno un IC>0.8 e IC>0.9 (fig. 3.9). La distribuzione dei siti che ne risulta è notevolmente diversa da quella mostrata nel

grafico di figura 3.8; i siti si distribuiscono in modo nettamente più omogeneo e non strettamente legato alla dimensione delle HTU.

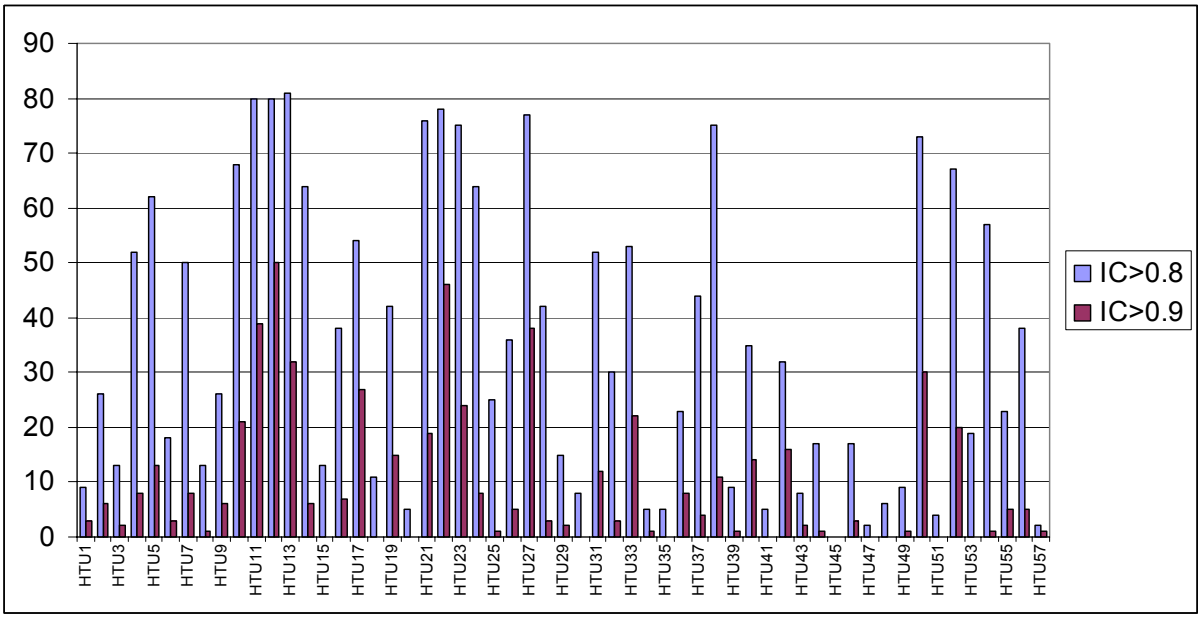


Figura 3.9. Siti con IC calcolato su tutte le HTU

A differenza del grafico di figura 3.8, che considerava per ogni HTU solo i siti in questa collocati, e in cui solo 18 HTU contenevano siti e di queste 13 avevano siti con IC>0.8, il grafico di figura 3.19 mostra che tutte le HTU (fatta eccezione per la HTU 45) hanno nel territorio europeo almeno un sito che ha un Indice di Coincidenza maggiore di 0.8 con le proprie caratteristiche. In particolare numerose HTU che non presentano alcun sito al loro interno, hanno invece un elevato numero di siti, al di fuori dei loro confini che sono ad esse altamente correlati. Si evidenziano in questo particolare caso le HTU 11, 13, 21, 23, 27, 38 e 50 che hanno più di 70 siti correlati con un IC>0.8. Le HTU 11 e 13 hanno un numero di siti con IC>0.8 almeno pari a quello dell'HTU 12 che essendo la più estesa del territorio, contiene un numero di siti interni pari a 40 e un numero totale di siti con IC>0.8 pari a 80. Puntando l'attenzione sulla penisola scandinava e sul nord della Francia che risultavano nell'analisi del precedente paragrafo prive di siti e quindi non rappresentate, emerge come in realtà le HTU 1 e 3 della Norvegia abbiano in totale 5 siti con IC>0.9 che le rappresentano, mentre la HTU 22 che è quella relativa all'area nel territorio francese ha ben 46 siti con un IC>0.9.

Le HTU che hanno più siti affini con le proprie caratteristiche non sono necessariamente le più grandi: la HTU 13 (32 siti correlati con IC>0.9) ha una estensione di quasi 200.000 Km^2 (tab.3.1), ma la HTU 11 (39 siti correlati con IC>0.9) misura meno di 20.000 Km^2.
Da questa distribuzione si deduce come la rappresentatività di un sito non si mostra perciò necessariamente legata alla sua posizione geografica. Esiste una rappresentatività ecofisiologica che permette a molte HTU di essere comunque rappresentate all'interno del territorio europeo.
L'alta correlazione tra siti e HTU non legati geograficamente tra loro, è comunque sicuramente legata all'ampiezza delle classi usate per la classificazione; un numero più elevato di classi avrebbe frammentato le HTU aumentando il numero di combinazioni sia per esse che per i dati provenienti dai siti, con il risultato di avere comunque valori più distribuiti e anche più bassi di IC.

3.6 Conclusioni

L'applicazione delle classiche tecniche di analisi GIS al dominio europeo e alla rete dei siti analizzata, ha permesso di individuare 57 aree di territorio, all'interno dei confini europei, che risultano omogenee per i parametri ecofisiologici che sono stati presi in considerazione (Homogeneous Territory Unit, HTU). Si distinguono la zona continentale europea e la penisola scandinava caratterizzate da estese HTU; si individua anche la zona mediterranea e la zona che comprende l'arco alpino dove le HTU risultano essere più piccole e frammentate a causa della elevata mosaicità del territorio (par. 3.5.1). Proiettando su queste aree omogenee i siti di monitoraggio, è stata valutata la loro distribuzione geografica, evidenziando la non uniforme distribuzione dei siti nelle aree di maggiori dimensioni e l'esistenza di estese HTU non contenenti alcun sito di monitoraggio (soprattutto nella penisola scandinava e nel nord della Francia); un eventuale ripensamento della rete dei siti dovrebbe prevedere l'inserimento di uno o più torri eddy in queste aree. Tuttavia l'area coperta dalle HTU che presentano siti al loro interno risulta essere il 74.78% dell'area totale e questo mostra una buona rappresentatività geografica dell'intera rete.
Per ogni sito considerato nella propria HTU è stato quindi calcolato l'Indice di Coincidenza (IC) a partire dalla sovrapposizione o meno delle classi in cui sono stati suddivisi i 5 parametri considerati. L'analisi dell'indice ha mostrato una buona correlazione tra i parametri misurati nei siti di monitoraggio e quelli derivati dai dati spazializzati: il 47.19% dei siti ha un

IC>0.9 con la HTU di appartenenza; se si considera una soglia per IC di 0.8, la percentuale dei siti con una buona correlazione sale all'83.15% (par. 3.5.2). Nel calcolo dell'Indice di Coincidenza, le variabili GPP (tra quelle classificate con 3 classi) e VPD (tra quelle classificate con 2 classi) sono quelle cha hanno presentato una variabilità maggiore, determinando i più bassi valori di IC. Le HTU con un più alto numero di siti altamente correlati sono localizzate nelle aree più omogenee dal punto di vista sia climatico che vegetazionale del territorio europeo.
Un ulteriore livello di analisi è stato aggiunto calcolando l'Indice di Coincidenza dei siti su tutte le HTU, a prescindere cioè dalla loro posizione geografica (par. 3.5.3). La distribuzione di siti in questo caso è nettamente diversa da quella mostrata considerando l'appartenenza geografica dei siti alle singole HTU. Tutte le HTU, salvo una, si relazionano almeno con un sito con un IC>0.8 e ben 47 HTU sulle 57 totali hanno un sito con un IC>0.9. Viene così evidenziata l'esistenza di una rappresentatività che esula dalla posizione geografica, ma che è determinata da analoghe condizioni dei parametri ecofisiologici che costituiscono il dominio considerato. In questo ambito aree che sono prive di siti risultano invece rappresentate dal punto di vista sia climatico che ecofisiologico (sempre nei termini del dominio studiato) da siti più o meno distanti geograficamente da loro. Una ipotetica ridistribuzione dei siti dovrebbe tenere conto anche di questa rappresentatività ecofisiologica piuttosto che limitarsi solo a pure considerazioni geografiche.

L'analisi mediante l'applicazione delle tecniche GIS condotta in questo studio presenta tuttavia dei limiti che influiscono in parte sui risultati ottenuti. In particolare la risoluzione dei dati spazializzati non ha permesso una più accurata classificazione per non frammentare estremamente il territorio europeo. Questo ha comportato un numero di classi per ogni variabile abbastanza piccolo, che si è tradotto in un IC poco distribuito nel range di 0 – 1.
Inoltre nelle aree più frammentate del territorio europeo il minor valore di IC potrebbe non essere effettivamente dovuto ad una non coincidenza del sito con l'HTU di appartenenza, ma alla effettiva difficoltà di poter relazionare dati misurati puntualmente con l'area in proporzione troppo estesa dei dati spazializzati. Un approfondimento quindi dell'analisi che preveda per le aree più frammentate l'uso di dati spazializzati con una maggiore risoluzione, permetterebbe di valutare in modo più preciso il significato dell'Indice di Coincidenza in queste stesse aree.

Capitolo 4

Analisi mediante applicazione di SOM

4.1 Introduzione

L'analisi condotta mediante classiche tecniche GIS descritta nel Capitolo 3 ha evidenziato come su scala continentale abbia un'importanza fondamentale il numero e la classificazione delle variabili nella individuazione di aree omogenee su cui confrontare la rete di monitoraggio.

Volendo ampliare il dominio della ricerca, inserendo non solo variabili ecofisiologiche e climatiche, ma anche la loro variabilità interannuale o stagionale, si determina una dimensionalità delle variabili tale che difficile diventa il trattamento dei dati spazializzati su questa scala con le classiche tecniche multilayer.

In genere la riduzione dei dati multivariati è normalmente condotta usando tecniche statistiche come l'Analisi delle Componenti Principali (PCA) o altre gerarchiche tecniche di Cluster Analisys (Jongman et al., 1995; Kaski, 1997). Ma spesso i tradizionali sistemi di simulazione basati sulle relazioni statistiche presentano una evidente difficoltà a rappresentare le complesse risposte degli organismi in ambito ecologico (Papale, Valentini, 2003). In questi anni si sono fatte strada così nuove tecniche investigative tra le quali i modelli neuronali emergono per varietà nelle applicazioni e per la capacità di simulare sistemi naturali.

Le reti neuronali si sono sviluppate inizialmente come modelli di neuroni biologici; esse hanno la capacità di apprendere dall'esperienza in un modo che nessun computer può elaborare e possono risolvere rapidamente complicati problemi computazionali. La ricerca sul loro sviluppo ha portato alla creazione di diversi tipi di reti neuronali utilizzate per risolvere problemi di memoria auto-associativa, di generalizzazione, ottimizzazione, riduzione dei dati, controllo e predizione in vari scenari e architetture (Lek, Guégan, 2000).

Le reti neuronali sono particolari modelli statistici basati su combinazioni lineari di funzioni appartenenti ad una classe avente opportune proprietà analitiche. Una rete neuronale può essere rappresentata mediante un grafo, dove gli esempi (pattern) sono costituiti da vettori di n-dimensioni assegnati ad alcuni nodi del grafo (nodi di ingresso e di uscita) e le

trasformazioni tra tali pattern sono ottenute mediante opportuni algoritmi. Una rete è quindi composta da una serie di unità elementari, i singoli neuroni (i nodi appunto), collegati tra loro grazie a connessioni pesate. Ciascun neurone riceve una serie di segnali, in risposta ai quali può attivarsi o meno e produce un segnale di output. Con il termine di *architettura* o *topologia* di una rete neuronale si indica la struttura della rete cioè il numero degli strati che si frappongono tra input e output, il numero di nodi in ciascuno di essi e le modalità con cui tali unità sono organizzate.

Le reti auto-organizzanti (Self Organizing Maps, SOM) fanno parte dei modelli neuronali e sono un ormai affermato strumento per la visualizzazione di dati multidimensionali.

Le reti SOM nascono all'inizio degli anni '80 (Kohonen, 1981) e sono un modello non parametrico che non richiede alcuna ipotesi di partenza circa la distribuzione dei dati presentando caratteristiche tipiche sia dei metodi di classificazione che di quelli di proiezione. Il loro vantaggio principale è la loro capacità di ricondurre spazi di n-dimensioni in una rappresentazione bidimensionale. Questo rende molto utile la visualizzazione della griglia bidimensionale per esplorare le relazioni tra le variabili e il possibile raccoglimento in cluster dei dati (Luyssaert et al., 2007).

Scopo di questa parte della ricerca è quella di applicare le SOM per individuare cluster omogenei all'interno del territorio europeo in rapporto al dominio costituito dalle variabili considerate e di valutare la posizione dei siti di monitoraggio Eddy nei cluster per dare una misura della loro rappresentatività anche in relazione alla variabilità interannuale e stagionale delle variabili. L'analisi, che viene condotta considerando separatamente ogni Plant Functional Type del dataset, si prefigge anche lo scopo di individuare possibili nuove distribuzioni dei siti della rete all'interno del territorio europeo, in modo tale da aumentare la sua rappresentatività nel dominio considerato.

4.2 Teoria di base delle SOM *(da Vistocco, 2002)*

4.2.1 Idea e topologia

Le reti neurali possono essere interpretate come una particolare classe di modelli matematico statistici rivolti alla risoluzione di differenti tipologie di problemi. E' possibile suddividerle, rispetto alla loro modalità di costruzione, in reti supervisionate e in reti non supervisionate. Nel primo caso al modello vengono applicati in fase di addestramento coppie di input e output: per ogni input da analizzare si presenta alla rete anche l'output desiderato e al termine

dell'addestramento il modello è usato per fini di previsione o decisionali. Le reti non supervisionate, come le reti SOM, non richiedono invece nella fase di addestramento la presenza dell'output: la rete analizza i dati di input alla ricerca di strutture e pattern e in base a queste si auto-organizza. I modelli di questo tipo sono in genere usati per classificare, organizzare e visualizzare grandi dataset. Per questo può essere fondamentale ai fini del risultato ottenuto, la fase di pre-processing dei dati; rientrano in questa fase sia la scelta di effettuare particolari trasformazioni delle variabili (come la normalizzazione), sia quella di fare assumere alle stesse un ruolo differente, per esempio introducendo un opportuno sistema di pesi (Pyle, 1999).

La caratteristica peculiare delle SOM sta nella modalità di apprendimento che può essere definito come non supervisionato ed insieme competitivo. Le unità o nodi della rete sono in genere disposti su una griglia bidimensionale. Ogni neurone è caratterizzato da un numero di connessioni pari al numero di input (ogni input viene infatti presentato a tutti i neuroni durante la fase di apprendimento) e da una serie di connessioni che lo legano ai neuroni che occupano una posizione vicina sulla griglia.

L'apprendimento è competitivo perché per ogni esempio presentato alla rete, i neuroni entrano in competizione tra loro: il neurone vincente sarà quello che presenta maggiori caratteristiche di somiglianza con lo specifico input. La regola di apprendimento determina il nodo più vicino e quindi il vincitore della competizione, come quello i cui pesi sono più vicini al pattern in esame. La vicinanza è misurata facendo ricorso solitamente alla classica distanza euclidea. L'apprendimento avviene solitamente in due fasi che sono la competizione e l'aggiornamento. Una volta stabilito il neurone vincente (competizione) i parametri dello stesso, cioè le connessioni che ad esso afferiscono, vengono aggiornati secondo una regola che lo spinge ancora di più verso l'input. Poiché però la griglia è caratterizzata da una rete di connessione tra i nodi, spostare un nodo verso una determinata direzione equivale a spostare in qualche modo tutti i neuroni ad esso collegati. Quindi l'apprendimento da competitivo si fa cooperativo nella seconda fase (aggiornamento): solitamente il modello procede all'aggiornamento non solo delle connessioni relative all'unità vincente, ma anche di quelle associate ai nodi a questa collegati. L'aggiornamento permette al nodo vincente di imparare a riconoscere gli input simili e poiché il processo coinvolge anche i nodi adiacenti, nodi vicini sulla griglia riconoscono input con caratteristiche simili. Gli esempi in input vengono presentati più volte (reiteramento della rete) aggiornando ad ogni passo i pesi dei nodi vincenti e di quelli a loro vicini. Inoltre l'aggiornamento localizzato fa in modo che solo

alcuni dei nodi della mappa siano interessati da valori anomali preservando gli altri e quindi il modello generale.

Al termine della fase di apprendimento la rete presenta quindi un certo grado di adattamento ai dati e può essere utilizzata sia sugli stessi dati usati durante l'addestramento (uso descrittivo) sia su nuovi dati (generalizzazione del modello).

Il confronto finale tra i vari neuroni, una volta terminata la fase di assegnazione degli input, permette di individuare classi di input simili (clustering). Le classi individuate potranno essere rappresentate ed individuate anche graficamente (proiezione del modello). Una SOM può quindi essere utilizzata su dati multidimensionali sia come strumento di classificazione, che come strumento di riduzione con una eventuale successiva visualizzazione.

Una SOM è in definitiva une rete neurale che usa un algoritmo di apprendimento non supervisionato. La rete si scompone di due soli strati di cui uno è lo strato di input e l'altro lo strato di output (fig. 4.1).

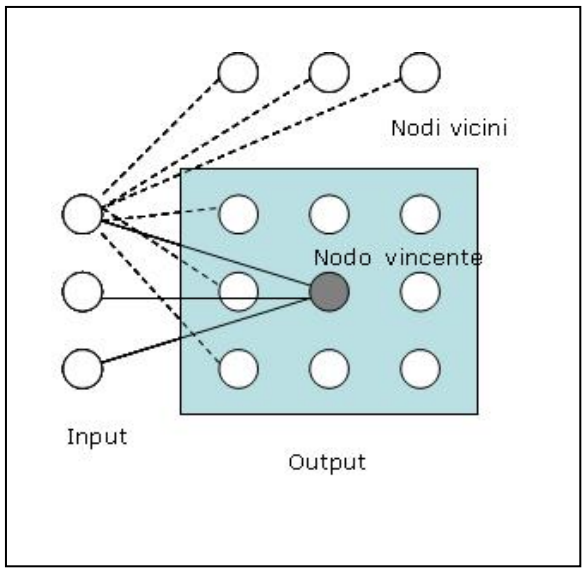


Figura 4.1. Rappresentazione grafica di una SOM

A differenza delle reti di tipo supervisionato questo tipo di rete non presenta quindi uno strato intermedio. Le unità dello strato di input sono pienamente connesse a quello dello strato di output, vale a dire che da ogni nodo dello strato di input parte una connessione verso ogni nodo dello strato di output. Le unità di output sono solitamente disposte su di una griglia bidimensionale e sono legate fra loro da una particolare funzione, detta *funzione di vicinanza* (funzione a bolla, funzione gaussiana, funzione gaussiana troncata o funzioni più complesse),

che per ciascun nodo stabilisce quali, tra i nodi della griglia, possono essere considerati ad esso vicini o adiacenti. E' possibile inoltre determinare vari livelli di vicinanza partendo dal nodo che interessa e spostandosi via via verso l'esterno.

Il numero di neuroni dello strato di input è determinato dalla dimensionalità dei dati di input (un nodo per ogni variabile del dataset) mentre il numero di neuroni dello strato di output è fissato dall'utente in base al livello di dettaglio desiderato e alla finalità di utilizzo del modello.

Esistono varie modalità differenti di organizzazione dei neuroni nello strato di output, ma le griglie bidimensionali sono quelle di più facile e immediata interpretazione. In questo studio le SOM sono visualizzate quindi con griglie bidimensionali a maglie esagonali (fig.4.2)

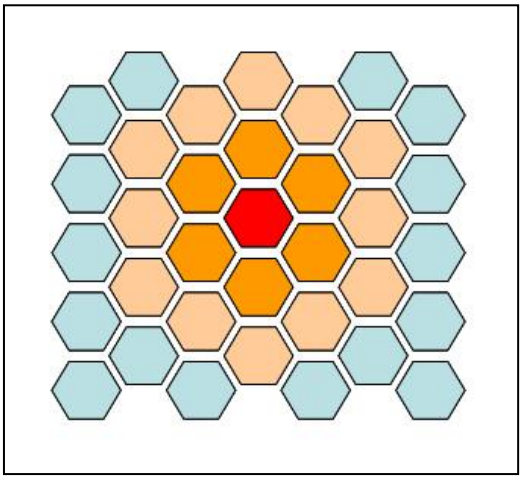


Figura 4.2. Tipologia di griglia bidimensionale a maglie esagonali con due livelli successivi di nodi adiacenti

4.2.2 La regola di apprendimento

Si consideri uno spazio a *p* dimensioni (cioè il caso in cui siano osservate *p* variabili di input). Una generica *i*-esima osservazione del dataset di input è costituita da un vettore a *p* dimensioni:

$$X_i = [\ X_{i1},\ldots,X_{ik},\ldots,X_{ip}\ ]$$

che corrisponde quindi ad una riga della matrice di input unità per variabili. Si indichi con D il numero di nodi della griglia di output. Ad ogni nodo *j*, (*j* = 1,…,*D*) della griglia è associato un vettore m_j di pesi che ha la stessa dimensionalità del vettore di input:

$$m_j = [\ m_{j1},...,m_{jk},...,m_{jp}\ ]$$

Il generico componente m_{jk} di questo vettore indica il peso associato alla connessione su cui viaggia il *k*-esimo segnale in input. Riferendosi ad un generico passo *t* del processo di apprendimento, che consiste di *T* differenti ripetizioni, è possibile indicizzare il vettore dei pesi in maniera da riferirsi ad un dato istante particolare della fase di apprendimento:

$$m_j(t) = [\ m_{j1}(t),...,m_{jk}(t),...,m_{jp}(t)\ ]$$

Il processo di addestramento consiste nel determinare il vettore $\Delta m_j(t)$ che deve essere sommato algebricamente al vettore $m_j(t)$ per aggiornare i pesi:

$$m_j(t+1) = m_j(t) + \Delta m_j(t) \qquad \text{eq. 4.1}$$

L'inizializzazione dei pesi può essere casuale: in questo caso i pesi sono inizializzati usando valori casuali opportunamente piccoli; oppure può essere decisa in base al campione e in questo caso i pesi sono inizializzati usando unità di input estratte casualmente dal dataset; può essere infine lineare quando invece si usano i due più grandi auto vettori dello spazio determinato dal dataset di input. L'algoritmo SOM si dimostra particolarmente robusto rispetto ai valori iniziali dei pesi per cui anche partendo da un insieme completamente casuale si riesce a raggiungere una soluzione stabile.
La fase di competizione consiste nell'estrazione casuale di una delle osservazioni di input (una riga della matrice) e nel determinare quale tra i *D* nodi della griglia può essere considerato il più vicino. Tale nodo viene designato come il vincitore della competizione. Si indichi con:

$$X_{(t)} = [\ X_{(t)1},...,X_{(t)k},...,X_{(t)p}\ ]$$

il vettore estratto casualmente dal dataset di input al tempo *t*, e con $m_w(t)$ il nodo vincente allo stesso ciclo di addestramento. Si avrà:

$$m_w(t) : \left\| X_{(t)} - m_w(t) \right\| = \min_{i=1,D} \{ \parallel X_{(t)} - m_i(t) \parallel \} \qquad \text{eq. 4.2}$$

Dove con $\parallel X_{(t)} - m_i(t) \parallel$ si è indicata una generica distanza tra i due vettori $X_{(t)}$ e $m_i(t)$.
Tale misura di distanza viene solitamente calcolata facendo ricorso alla distanza euclidea:

$$\left\| X_{(t)} - m_w(t) \right\| = \sqrt{\sum_{i=1}^{p} \left(X_{(t)k} - m_{ik}(t) \right)^2} \qquad \text{eq. 4.3}$$

Una volta identificato il nodo vincente è necessario procedere all'aggiustamento dei pesi ad esso afferenti. L'aggiornamento viene effettuato in maniera da avvicinare ulteriormente il nodo vincente $m_w(t)$ a $X_{(t)}$: se in un passo successivo dell'addestramento verrà ripresentato alla rete lo stesso esempio o un esempio con caratteristiche simili, il nodo tenderà sempre a vincere la competizione; in questo modo ogni nodo della rete impara a riconoscere input simili. Il valore di $\Delta m_j(t)$ viene solitamente determinato in maniera proporzionale alla distanza tra $X_{(t)}$ e $m_w(t)$:

$$m_w(t+1) = m_w(t) + \alpha(t) \left[X_{(t)} - m_w(t) \right] \qquad \text{eq. 4.4}$$

Il fattore di proporzionalità $\alpha(t)$, compreso tra 0 e 1, è il *tasso di apprendimento*. Anche questo risulta essere indicizzato a *t* e solitamente decresce con l'aumentare di *t*. Nelle prime epoche di addestramento il valore è vicino all'unità provocando un aggiustamento sostanziale di $m_w(t)$ a $X_{(t)}$: tale valore tende invece a zero verso le fasi conclusive dell'addestramento in maniera da limitare gli spostamenti dei pesi una volta che la rete ha assunto una configurazione abbastanza stabile.
Come già descritto le SOM sono caratterizzate da un aggiornamento non solo del nodo vincitore della competizione ma anche dei nodi vicini. Nel caso più semplice i legami tra i

nodi sono stabiliti in maniera fissa. La funzione di vicinanza può assumere una forma più o meno complessa. La forma delle funzioni di vicinanza è tale che i nodi più vicini al nodo vincente $m_w(t)$ sono aggiornati in misura maggiore rispetto a quelli più lontani. Un esempio tipico di funzione di vicinanza che permette questo aggiornamento inversamente proporzionale alla distanza del nodo rispetto al nodo vincente è la funzione gaussiana:

$$h_{ij} = \exp\left\{\frac{-(i-j)^2}{2\sigma^2}\right\} \qquad \text{eq. 4.5}$$

In questo caso h_{ij} è funzione degli indici i e j e non della distanza tra X e gli m_i: questo provoca un aggiornamento di tutti i pesi della griglia di output per ogni passo del processo di addestramento.
Un'ulteriore complicazione si può introdurre indicizzando anche la funzione di vicinanza al passo specifico t del processo di apprendimento:

$$m_i(t+1) = m\ (t) + \alpha(t)h_{wi}(t)\left[\ x_{(t)} - m_i(t)\ \right] \qquad \text{eq. 4.6}$$

In questo caso si utilizza in genere una funzione più ampia nella parte iniziale dell'addestramento, che viene poi ristretta sempre più attorno al nodo vincente al crescere delle iterazioni. La presenza della funzione di vicinanza, sia essa determinata da una semplice relazione di appartenenza o da una forma funzionale più complessa, è caratteristica di questo tipo di modelli ed è il motivo per cui, riguardo alla seconda fase del processo di addestramento, si parla di apprendimento cooperativo.
La combinazione dell'indicizzazione sia del tasso di apprendimento che della funzione di vicinanza determina una sorta di suddivisione dell'apprendimento in due momenti successivi: inizialmente si ha un adattamento grossolano del modello ai dati per poi raggiungere successivamente un livello di dettaglio maggiore. In altre parole è come se nella prima fase si scegliessero le direzioni principali di adattamento, cioè le relazioni globali presenti nel dataset, per poi cercare di catturare nella seconda fase relazioni locali.

Questa regola di apprendimento è quella originariamente proposta da Kohonen. Successivamente in letteratura sono state proposte numerose varianti rivolte migliorare la performance e l'efficienza.

Una di queste varianti è quella applicata in questo studio ed è la *versione batch*. Si tratta di un algoritmo significativamente più veloce che non richiede la specificazione del tasso di apprendimento $\alpha(t)$. La differenza è determinata dalla modalità di presentazione degli esempi alla rete: invece di utilizzare un singolo vettore per ogni passo $t = 1,\ldots,T$, procede all'aggiustamento dei pesi solo dopo la presentazione alla rete dell'intera matrice. Per ogni nodo m_i della mappa si memorizza una lista contenente tutte le osservazioni in input X_i in cui il nodo è risultato vincente nella fase di competizione. Al termine della presentazione di tutti gli esempi, la regola di addestramento determina i nuovi pesi come media ponderata dei dati in input assegnati al nodo o ad uno dei suoi vicini, usando come peso il valore assunto dalla funzione di vicinanza:

$$m_i(t+1) = \sum_{k=1}^{n} \frac{h_{wi}(t)x_k}{h_{wi}(t)} \qquad \text{eq. 4.7}$$

4.2.3 La visualizzazione

L'assegnazione delle unità di input ai vari nodi della rete avviene in maniera che ad unità vicine vengono assegnati input simili: questo consente un utilizzo della rete come strumento di visualizzazione per esplorare pattern presenti in un dataset.

Il più comune grafico ottenibile con una SOM è riportato in figura 4.3. Questa rappresentazione è costruita a partire da una matrice di distanze di ogni unità della griglia di output rispetto ai nodi vicini. Tali distanze sono rappresentate su una scala di colori; i valori effettivi delle distanze tra i nodi sono i bordi che separano le varie unità mentre il colore di riempimento è ottenuto usando la mediana delle distanze rispetto alle unità adiacenti.

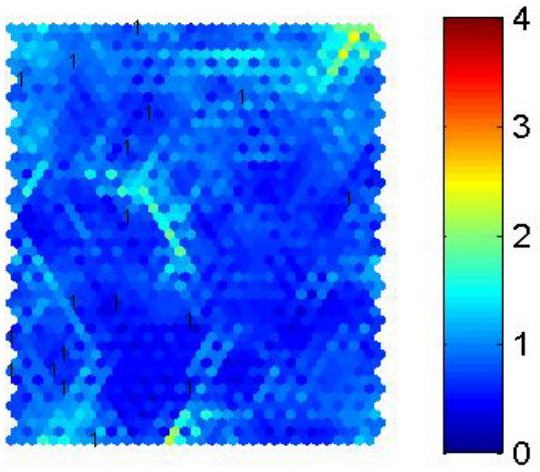


Figura 4.3. Rappresentazione delle matrice delle distanze

Distanze più grandi o più piccole tra le unità corrispondono a distanze più grandi o più piccole dei vettori dei pesi ad essi afferenti.

La matrice delle distanze è un comune mezzo usato per dividere in cluster una mappa SOM (Ultsch, Siemon, 1990). Per individuare i cluster è possibile guardare alla uniformità della distribuzione dei colori sulla mappa. Il numero dei cluster desiderato può essere stabilito sia visivamente sia stabilendo un livello desiderato di similarità sulla scala. Quando i cluster vengono stabiliti visivamente però, il risultato ottenuto risulta spesso soggettivo e non necessariamente uguale per ogni operatore; normalmente quindi vengono usati algoritmi automatici (Luyssaert et al., 2004).

Generalmente la matrice delle distanze è utilizzata per identificare il centro dei cluster, le altre celle della mappa sono poi assegnate al cluster il cui centro è più vicino (Vellido, 1999). Un ulteriore implementazione di questo modello di clusterig è determinata dalla individuazione di cluster di base che possono essere considerati dei super-cluster all'interno dei quali si trovano diversi sotto-cluster. Nell'analisi dei dati questo permette di investigarli in diversi livelli di dettaglio. Per costruire una gerarchia di cluster da ciascun set di cluster di partenza, può essere usato un algoritmo per agglomerare i cluster stessi; vengono così costruiti "alberi binari" che contengono extra-cluster che devono essere tagliati fuori in una specifica "fase di taglio". I dettagli di questo algoritmo si trovano in Vesanto 2002. I cluster individuati a partire dalla matrice delle distanze possono essere visualizzati con una mappa delle stesse dimensioni (fig. 4.4)

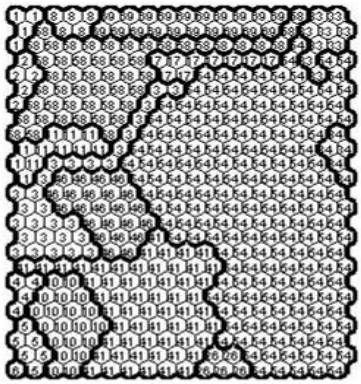


Figura 4.4. Rappresentazione della distribuzione dei cluster

La griglia bidimensionale che caratterizza le SOM può essere infine utilizzata anche per rappresentare singolarmente ogni variabile (fig. 4.5). Questo tipo di rappresentazione è costruita attraverso quelle che Kohonen definisce le componenti del piano. Ogni componente può essere vista come una fetta della mappa finale (Kohonen, 2001) e consiste nella proiezione di ogni singola variabile sulla mappa. L'esame della distribuzione dei colori permette di visualizzare l'andamento di quella particolare variabile. I nodi delle griglie della matrice delle differenze e delle singole componenti del piano sono collegati dalla posizione occupata: una cella cioè occupa in tutte le matrici la medesima posizione. La lettura incrociata della matrice delle distanze e delle variabili permette di caratterizzare i cluster ottenuti mentre il confronto tra l'andamento delle variabili permette di leggere le eventuali correlazioni tra esse.

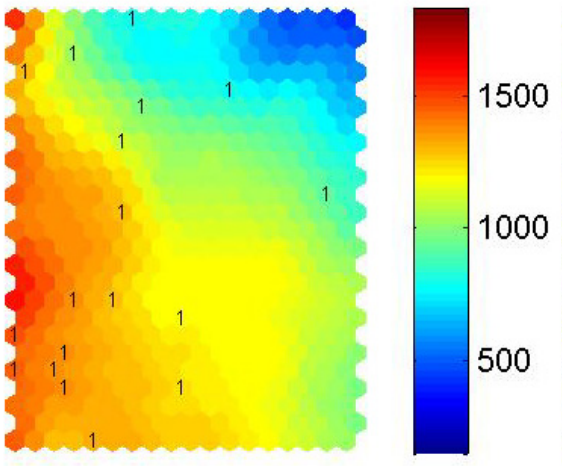


Figura 4.5. Rappresentazione di una componente del piano

La scelta dell'ampiezza della griglia, cioè il numero dei neuroni che la compongono deve essere legata a considerazioni sia sul livello di dettaglio che si vuole ottenere, sia sull'utilizzo che si vuole fare del modello. Una mappa di dimensioni maggiori, se da un lato permette un grado maggiore di dettaglio, dall'altro può creare problemi di overfitting, mentre una mappa di piccole dimensioni offre una maggiore possibilità di generalizzazione ma ha un dettaglio di analisi più grossolano.

4.3 Metodologia applicata

4.3.1 Il dominio

In questa parte della ricerca tutte le analisi sono state condotte considerando separatamente le 5 Plant Functional Type (PFT), tre di tipo forestale (EBF, DBF, ENF), le praterie (G) e le colture (C), così come descritto nel paragrafo 2.4.
Il dominio spaziale è quindi rappresentato dai pixel che nei dati spazializzati contengono uno specifico PFT (il dominio spaziale è quindi diverso per ogni PFT). Il pixel è stato assegnato ad un determinato Plant Functional Type se questo è presente al suo interno per almeno il 3% della sua area. In un dato pixel si possono trovare più PFT diversi e quindi questo farà parte del dominio spaziale di ogni PFT presente, in misura proporzionale alla sua percentuale di copertura. La mappa del Land Cover che è stata utilizzata per determinare le percentuali di copertura è fornita dal sistema MODIS (vedi tabella 2.3).

Per ognuno dei PFT sono state prese in considerazione 7 variabili ecofisiologiche e climatiche:

GPP	gC m^{-2} y^{-1}
NEE	gC m^{-2} y^{-1}
Precipitazioni annue (Precip.)	mm y^{-1}
Radiazione globale (Rad)	W m^{-2}
Temperatura media (Temp)	°C
Vapour Pressure Deficit (VPD)	KPa
Filtered PAR (fPAR)	%

Per ognuna delle variabili sono poi state prese in considerazione la media annuale e la variabilità interannuale e stagionale calcolate secondo le equazioni:

Media interannuale: $m = \bar{x}\left(\sum annuale\right)$ per GPP, NEE, Precip eq. 4.8

$m = \bar{x}\left(\overline{X}annuale\right)$ per Rad, Temp, VPD, fpar eq. 4.9

Variabilità interannuale $va = \frac{SD\left(\sum annuale\right)}{\bar{x}\left(\sum annuale\right)}$ per GPP, NEE, Precip eq. 4.10

$va = \frac{SD\left(\overline{X}annuale\right)}{\left(\overline{X}annuale\right)}$ per Rad, Temp, VPD, fpar eq. 4.11

Variabilità stagionale $vs = \frac{SD\left(\sum mensile\right)}{\bar{x}\left(\sum mensile\right)}$ per GPP, NEE, Precip eq. 4.12

$vs = \frac{SD\left(\overline{X}mensile\right)}{\left(\overline{X}mensile\right)}$ per Rad, Temp, VPD, fpar eq. 4.13

Il dominio basato sui parametri su cui valutare la rappresentatività dei siti di monitoraggio è stato così descritto da un totale di 21 variabili (dominio a 21 dimensioni).

4.3.2 La struttura SOM applicata

Le 21 variabili del dominio sono state sottomesse alla SOM. Essendo variabili espresse in unità di misura diverse, esse sono state dapprima standardizzate con una procedura di normalizzazione (media nulla e varianza unitaria).

La SOM è stata trainata usando l'algoritmo *batch* (vedi paragrafo 4.2.2): l'inizializzazione dei pesi è di tipo lineare e la rete è stata reiterata due volte nella prima fase di addestramento e 5 volte nella seconda fase.

La funzione di vicinanza applicata ai nodi della matrice bidimensionale è di tipo gaussiano.

Il numero dei nodi delle griglie di output è stato scelto per ogni PFT in base al numero dei pixel che nel land cover contengono almeno il 3% di copertura del PFT stesso.

La matrice delle distanze è stata suddivisa in cluster usando l'algoritmo "agglomerativo" (Han, Kamber, 2001) con "potatura" degli extra-cluster descritto in Vesanto 2002. Questo ha generato per ogni PFT un numero di super-cluster contenenti diversi sub-cluster: tra questi

super-cluster è stato scelto quello che contiene il numero massimo di sub-cluster che è possibile rappresentare con i siti della rete, quello cioè che contiene un numero di sub-cluster pari o immediatamente inferiore al numero dei siti dello specifico PFT (tab.4.1). Un numero maggiore di cluster rispetto al numero dei siti avrebbe infatti come conseguenza certa la non rappresentatività della rete, che è legata alla capacità dei siti stessi di coprire nel modo più omogeneo possibile i cluster presenti.

PFT	n° celle	n° siti	n° cluster
EBF	1428	10	10
DBF	7809	19	19
ENF	8330	18	18
G	3650	24	18
C	7858	18	18

Tabella 4.1. Grandezza delle mappe e numero dei cluster delle PFT

Nella matrice delle differenze e nelle matrici delle componenti del piano vengono infine proiettati i siti per valutare la loro distribuzione sul dominio determinato dallo spazio in 21 dimensioni.

Una volta individuati i cluster la rete SOM viene infine usata anche per calcolare il centro di ogni cluster. Questo si ottiene calcolando la media dei valori delle celle di un cluster ed individuando la cella che ha il valore più vicino a questa media, che sarà quello che ha la distanza quadratica media minore. Una volta individuato il centro (che non è quindi geometrico) del cluster, viene calcolata la distanza euclidea nello spazio a 21-D di ogni pixel dal centro del cluster.

4.4 Risultati e discussioni

4.4.1 Le mappe SOM

Per ognuno dei 5 Plant Functional Type sono state prodotte le mappe delle matrici delle differenze (Umatrix) e delle 21 componenti del piano (fig. 4.6 a-e). La diversa dimensione delle griglie è dovuta al diverso numero di pixel che nel territorio europeo sono occupate dallo specifico PFT (tab. 4.1).

Nelle mappe vengono rappresentati anche i cluster che sono generati in base alla matrice delle distanze Umatrix; un cluster è determinato da un insieme di celle che sono vicine tra loro e con un alto indice di similarità, mentre i confini dei cluster sono segnati dalle celle "più lontane" (valori più alti sulla scala delle distanze).

Il numero 1 o 2 che si trova all'interno delle mappe indica la posizione che nel dominio occupano i siti della rete per ogni PFT. Più i siti sono distribuiti uniformemente nei cluster, più la rete è rappresentativa. Quando i siti sono concentrati in soli pochi cluster, la rappresentatività diminuisce.

In generale la bassa variabilità che si nota in alcune mappe relative alle variabilità stagionali e annuali non è dovuta ad una effettiva omogeneità, ma alla scala con cui vengono rappresentate: alcuni valori estremi presenti per esempio nelle mappe relative alle Grassland (fig. 4.6 d) per la variabilità annuale della temperatura e per la variabilità stagionale della NEE (angolo in basso a destra) hanno determinato uno spostamento della scala verso alti valori con una conseguente contrazione della variabilità per i valori più bassi. Questo rappresenta solo un limite visivo dell'analisi in quanto nella determinazione dei cluster e della matrice delle differenze la variabilità stagionale e annuale è implicita nei dati e quindi compresa. In generale la variabilità annuale è comunque quella che presenta la più alta omogeneità in tutti i PFT, fatta eccezione per la variabilità annuale di NEE che presenta una distribuzione meno omogenea. La variabilità stagionale presenta invece all'interno di ogni PFT una distribuzione maggiore lungo la scala dei valori rappresentati.

La Umatrix relativa alle EBF (fig. 4.6a) sembra guidata principalmente dalla NEE_m, dalla Rad_m, dalla Rad_{vs} e dalla VPD_{vs} che hanno una variabilità maggiore e quindi pesano in modo più rilevante sulla costituzione dei cluster, mentre la GPP risulta essere meno determinante sulla guida della Umatrix. In generale tuttavia le singole componenti del piano mostrano una variabilità maggiore rispetto al numero dei cluster che è determinato dal numero dei siti presenti nella rete di monitoraggio. Questo porta a considerare che il numero delle torri per le EBF potrebbe non essere sufficiente per rappresentare la variabilità europea per il dominio considerato. I siti inoltre sono concentrati in due soli cluster, e anche se uno di questi è quello con maggiori dimensioni, anche la sola diversa distribuzione dei siti porterebbe ad una più alta rappresentatività.

Per le DBF invece le variabili che guidano la matrice delle differenze sono molteplici: GPP, NEE, Rad, VPD e Prec.(fig. 4.6b). Ciò significa che le singole variabili hanno un peso paragonabile le une alle altre e tutte contribuiscono alla determinazione dei cluster senza che

nessuna risulti preponderante. Inoltre le variazioni stagionali e annuali delle variabili in generale non risultano avere un peso nella guida della Umatrix. I cluster hanno dimensioni comparabili tra loro ma i siti si distribuiscono all'interno dei cluster in modo disomogeneo e molti cluster anche di dimensioni relativamente grandi non risultano rappresentati.

La matrice delle differenze delle ENF (fig. 4.6c) è determinata anch'essa da tutte le componenti del piano senza che nessuna di esse sia preponderante sulla guida della Umatrix stessa. Si notano dei valori estremi nella componente del piano NEE_{va} che guidano la Umatrix nella parte centrale a sinistra, determinando per quei nodi una grande distanza che li contraddistingue dagli altri. Questi nodi non hanno comunque tra loro caratteristiche così simili da farli raggruppare in un singolo cluster. È invece individuato un cluster centrale di grandi dimensioni in cui si trova la quasi totalità dei siti, mentre i cluster più piccoli si trovano ai bordi della mappa e sono privi di siti. Anche in questo caso una diversa distribuzione dei siti porterebbe ad un grado di rappresentatività più alto.

EBF **U – matrix** **Cluster**

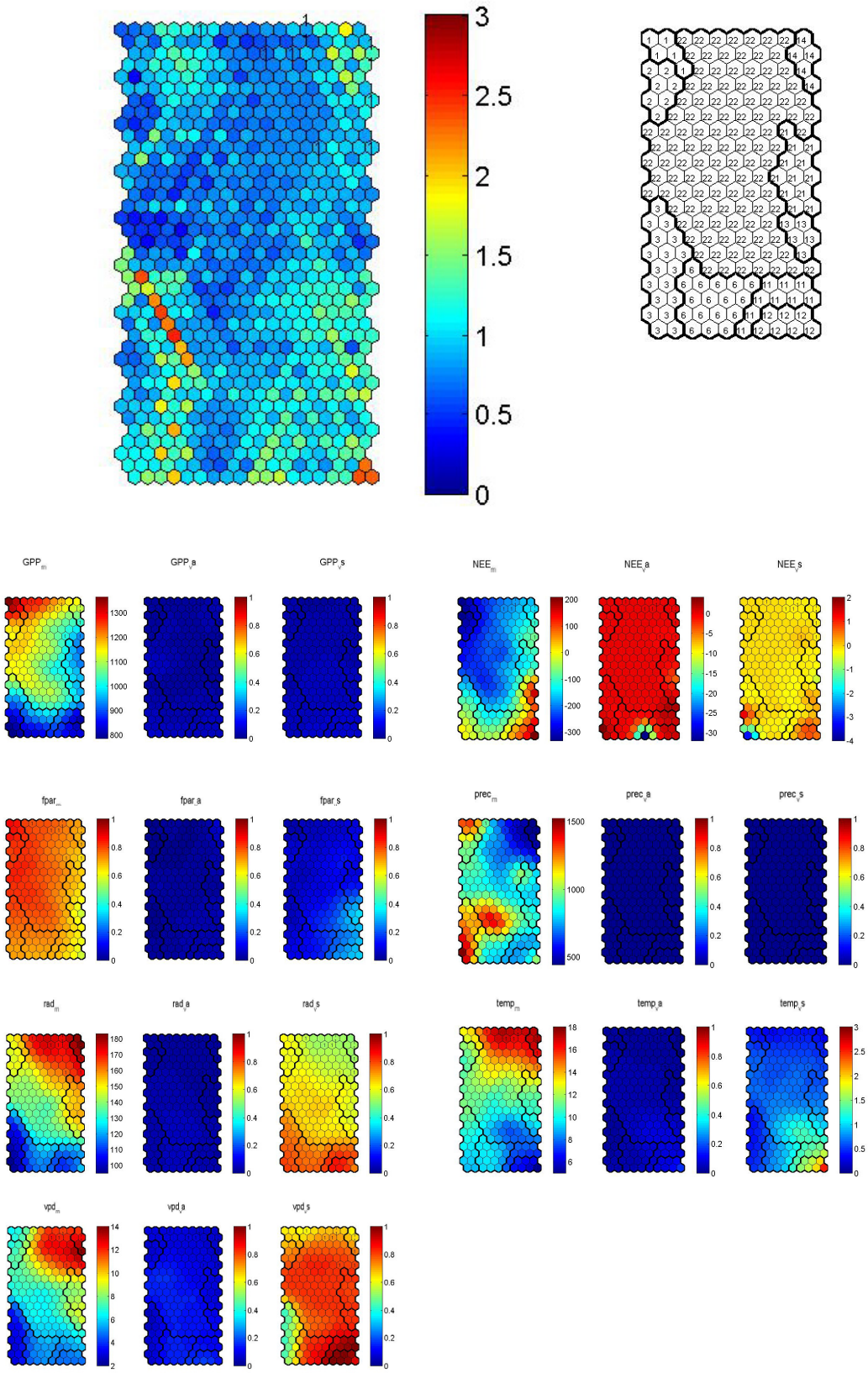


Figura 4.6a. Mappe SOM delle differenze e di ciascuna componente del piano per EBF

DBF **U – matrix** **Cluster**

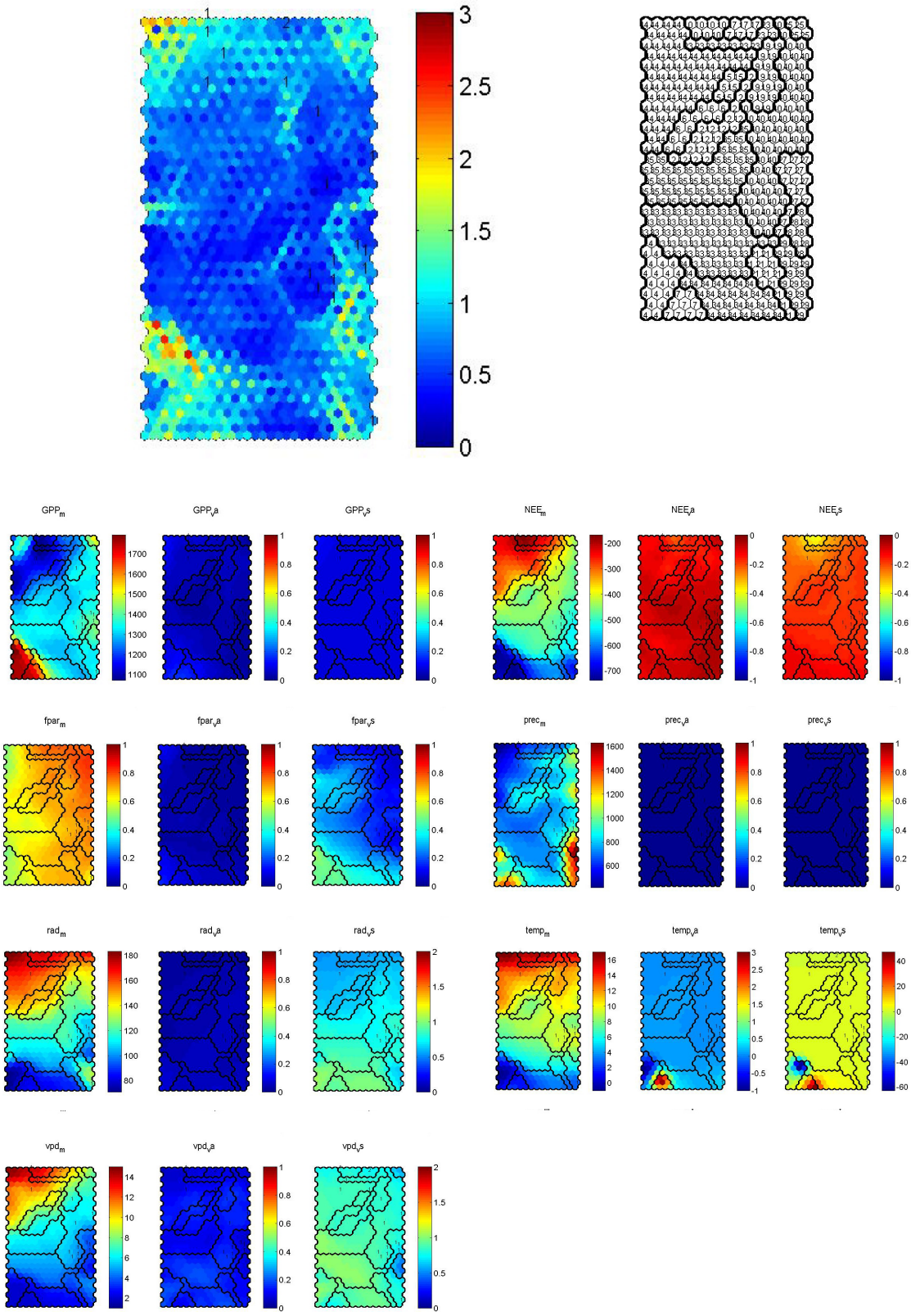


Figura 4.6b. Mappe SOM delle differenze e di ciascuna componente del piano per DBF

ENF **U – matrix** **Cluster**

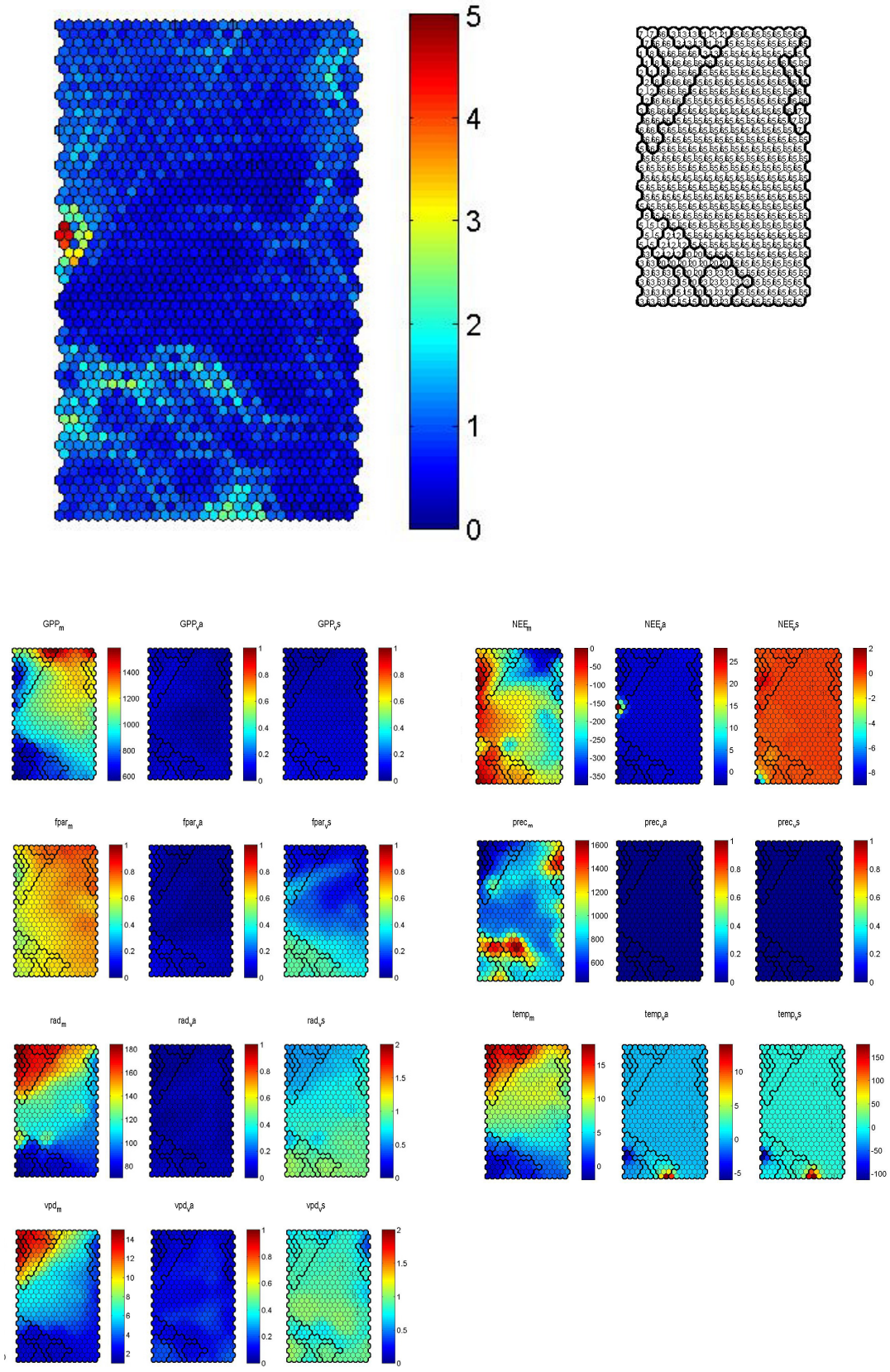


Figura 4.6c. Mappe SOM delle differenze e di ciascuna componente del piano per ENF

Gras **U – matrix** **Cluster**

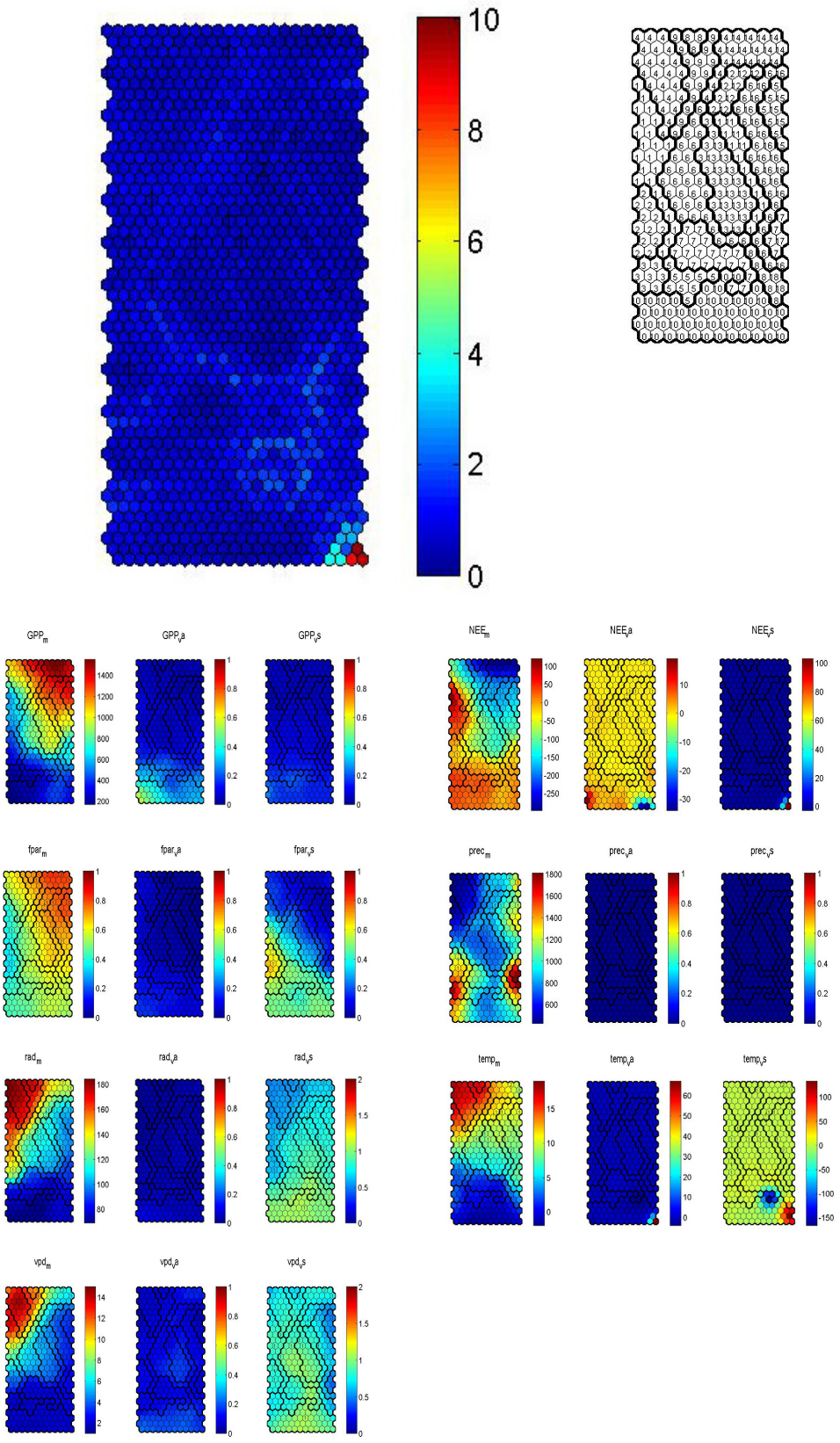


Figura 4.6d. Mappe SOM delle differenze e di ciascuna componente del piano per G

Crop **U – matrix** **Cluster**

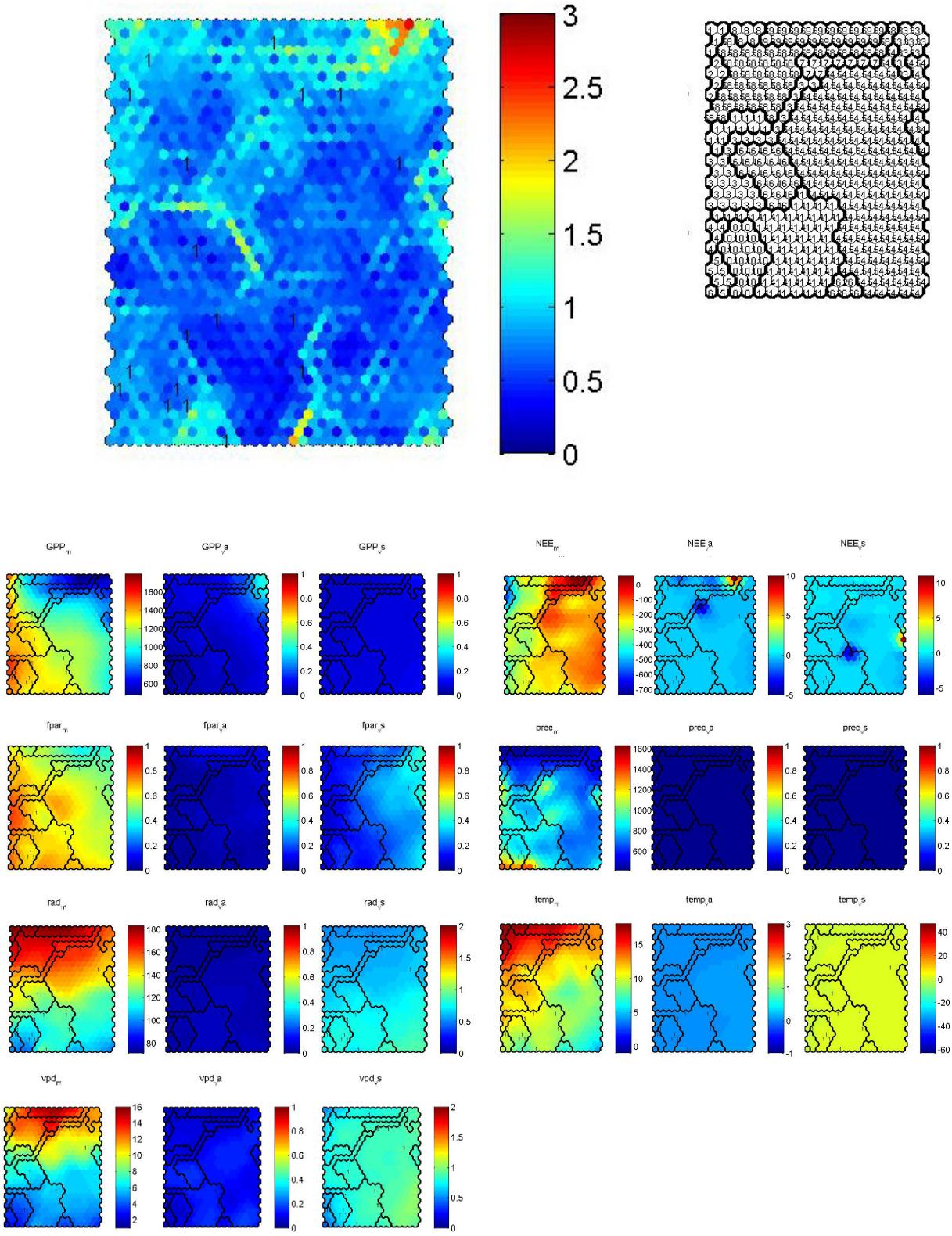


Figura 4.6e. Mappe SOM delle differenze e di ciascuna componente del piano per C

Per quanto riguarda le praterie (Grassland, fig. 4.6d) la Umatrix viene guidata soprattutto dalla NEE_m, dalla media delle precipitazioni ($Prec_m$) e dalla variabilità stagionale della VPD ($VPD_{vs.}$) I cluster che ne risultano sono abbastanza omogenei per dimensioni e i siti presentano una buona distribuzione sulle mappe e sui cluster. Alcuni valori anomali di NEE_{vs} e di $Temp._{vs}$ (nodi in basso a destra nelle due componenti del piano) conducono la matrice delle distanze ad avere nodi molti distanti tra loro nella stessa posizione, ma anche in questo caso la loro variabilità e il loro esiguo numero non determinano l'individuazione di un cluster specifico.

Nelle SOM relative alle colture (Cropland, fig. 4.6d), si individua un cluster di grandi dimensioni posizionato sulla destra delle mappe; per la determinazione di questo cluster nella matrice delle differenze la componente NEE_m risulta la variabile guida. Rad_m, $Temp_m$ e VPD_m guidano invece la Umatrix nell'identificazione dei cluster minori. GPP e precipitazioni non sembrano guidare in modo determinante la creazione della matrice delle differenze, così come tutte le componenti del piano relative alla variabilità annuale e stagionale. I siti sono distribuiti in modo abbastanza omogeneo all'interno dei cluster creati dalle SOM.

Confrontando quindi le mappe SOM per i tre PFT forestali, si evidenzia come in generale a parità di cluster generati, esse abbiano siti maggiormente concentrati solo in alcuni cluster, rispetto alle praterie e alle colture dove i siti sono più omogeneamente distribuiti, e questo è indice di una rappresentatività generalmente più bassa.

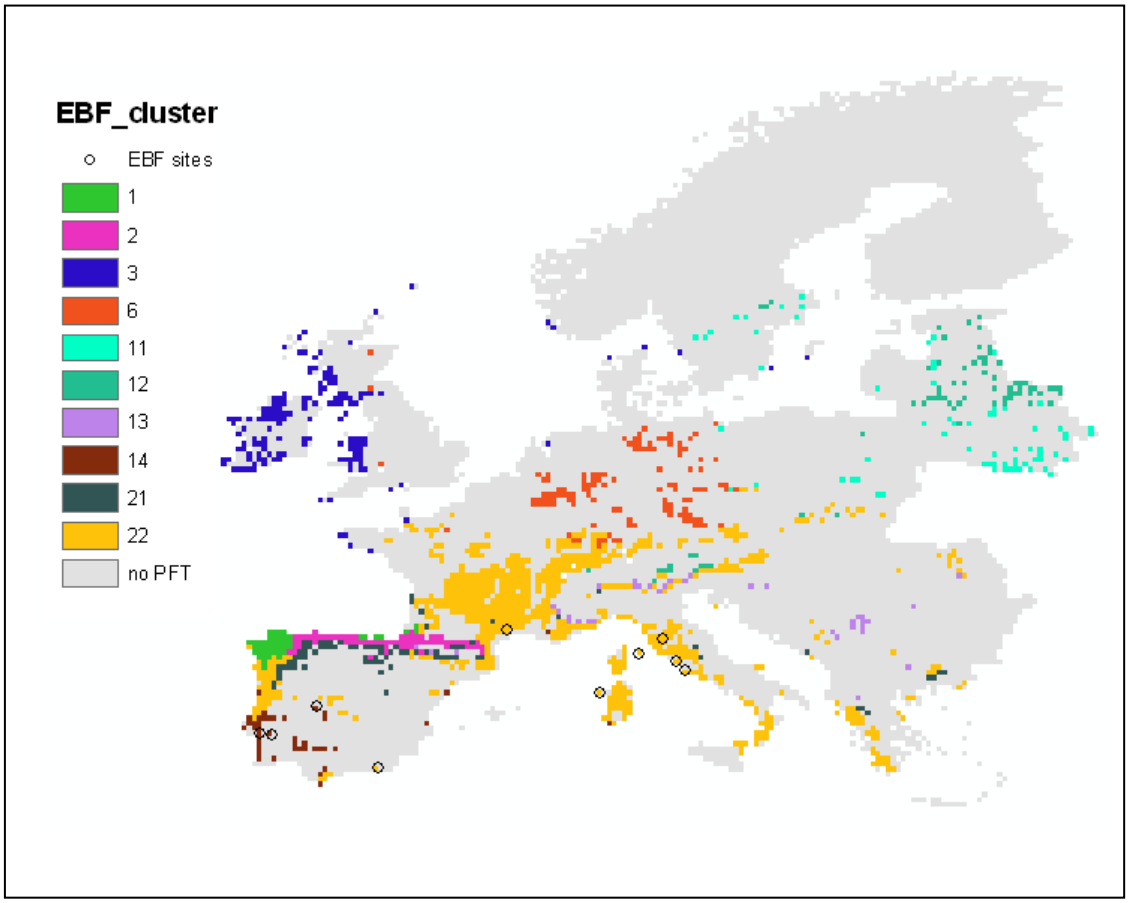
4.4.2 La distribuzione geografica

Per poter meglio visualizzare come si dispongono i siti all'interno dei cluster determinati con l'applicazione delle SOM, questi ultimi sono stati riportati su scala geografica all'interno del dominio spaziale europeo.

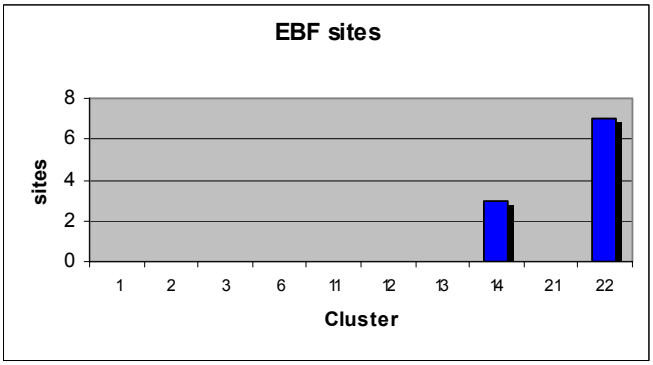
Per ciascun PFT viene quindi ricostruita la mappa geografica dei cluster e dei siti; per ogni cluster vengono quindi indicati con più chiarezza rispetto alle mappe SOM, il numero e la posizione dei siti contenuti (fig. 4.7 a-e).

Per poter dare inoltre una misura quantitativa relativa alla rappresentatività dei siti appartenenti ai singoli PFT nel loro specifico dominio, dalla mappa vengono calcolate le aree relative a ciascun cluster (pesate per la % di copertura dello specifico PFT presente nel pixel) e l'area totale coperta da tutti i cluster.

Sommando infine le aree dei cluster che contengono almeno un sito, si determina la percentuale di copertura delle aree contraddistinte dalla presenza dei siti stessi.

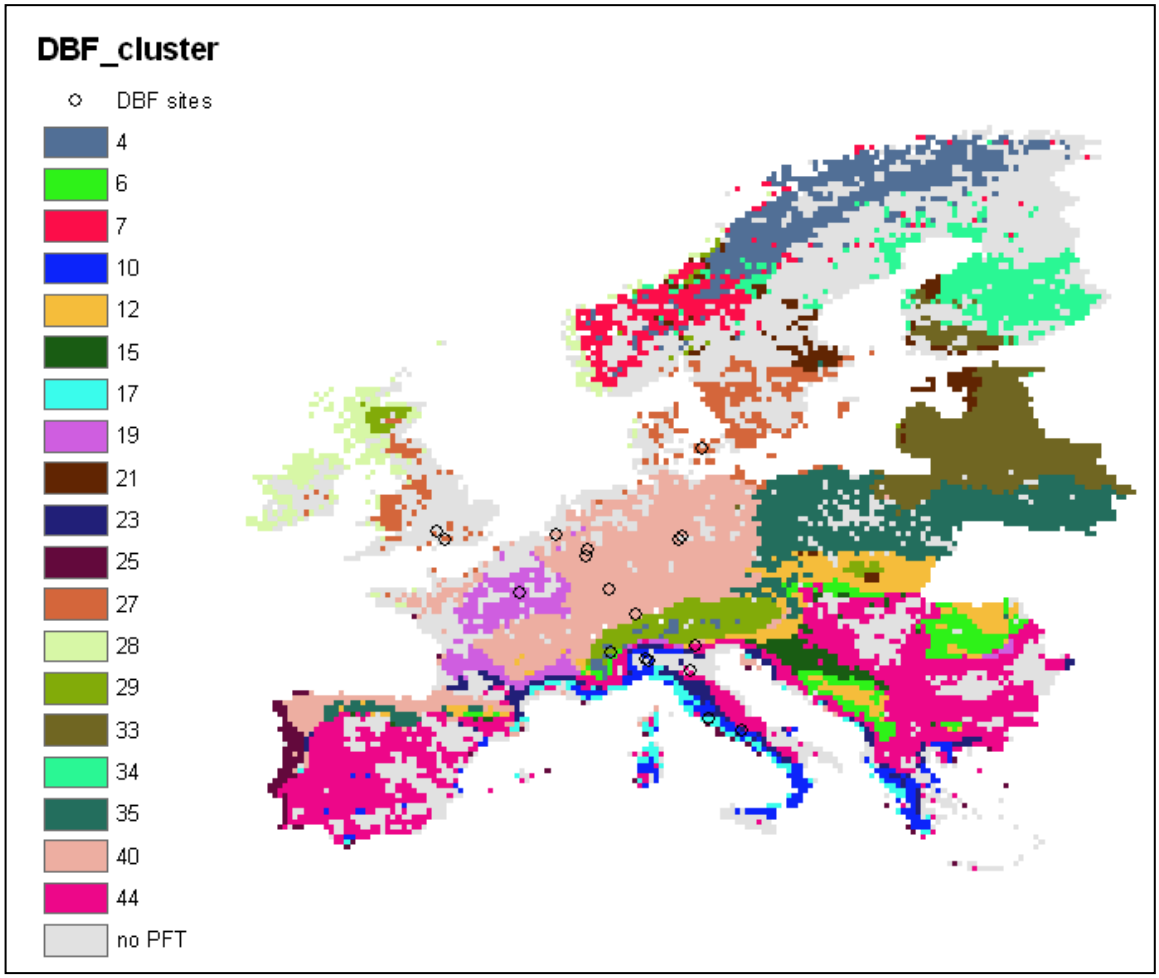


cluster	n°sites	cluster area (Km2)
1	0	4760
2	0	3022
3	0	1713
6	0	987
11	0	447
12	0	471
13	0	303
14	3	857
21	0	1003
22	7	16661
	Total Area	30224

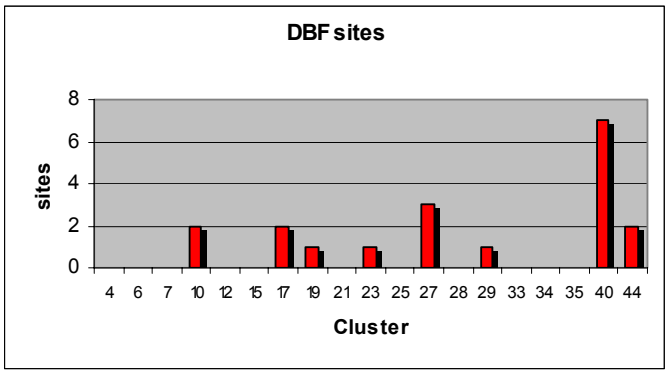


% PFT Area **57.96**

Figura 4.7a. Distribuzione geografica dei cluster e dei siti per EBF

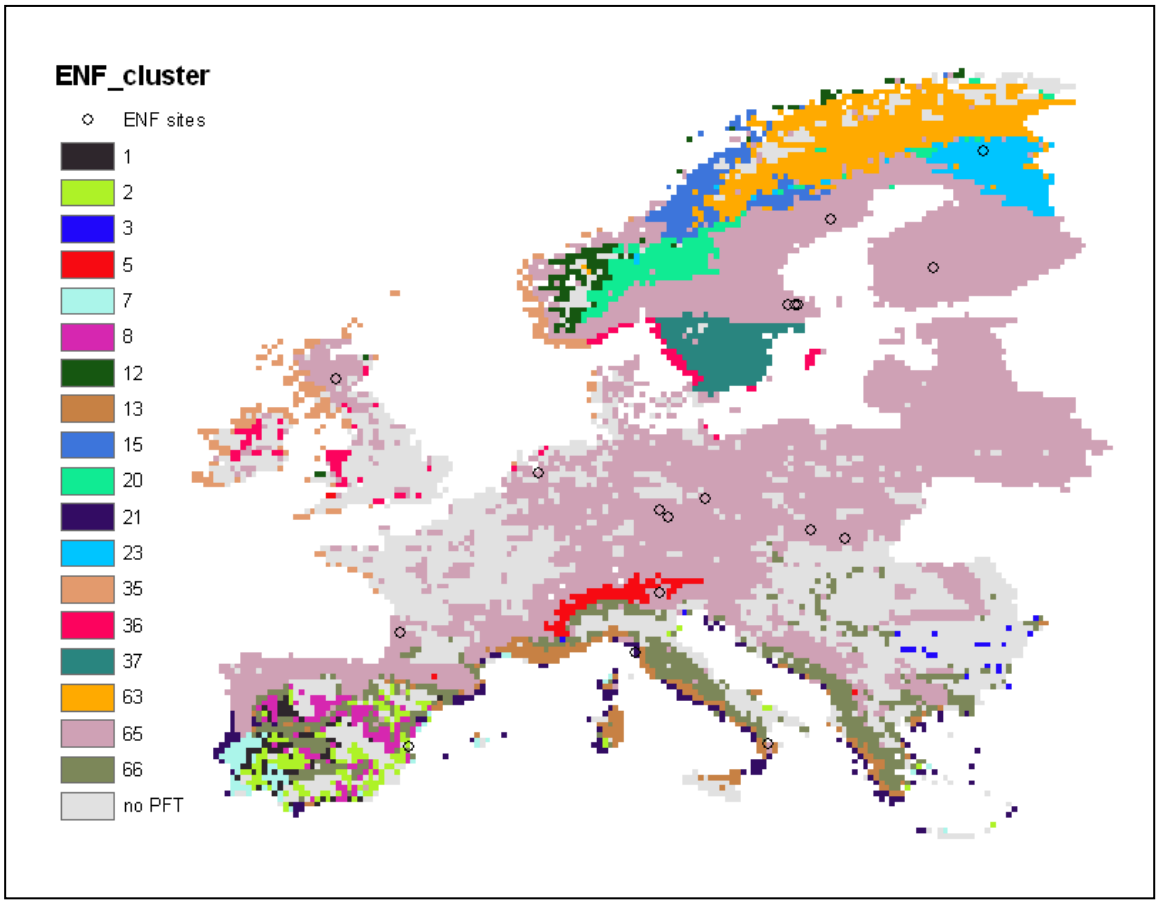


cluster	n°sites	cluster area (Km2)
4	0	21786
6	0	39455
7	0	7049
10	2	21593
12	0	57153
15	0	28694
17	2	11569
19	1	21665
21	0	6696
23	1	22891
25	0	9514
27	3	8414
28	0	6306
29	1	30094
33	0	42587
34	0	8961
35	0	41228
40	7	130823
44	2	140559
	Total Area	657037

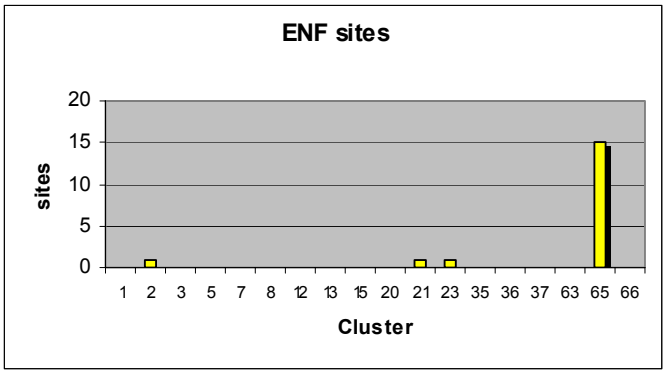


% PFT Area **58.99**

Figura 4.7b. Distribuzione geografica dei cluster e dei siti per DBF

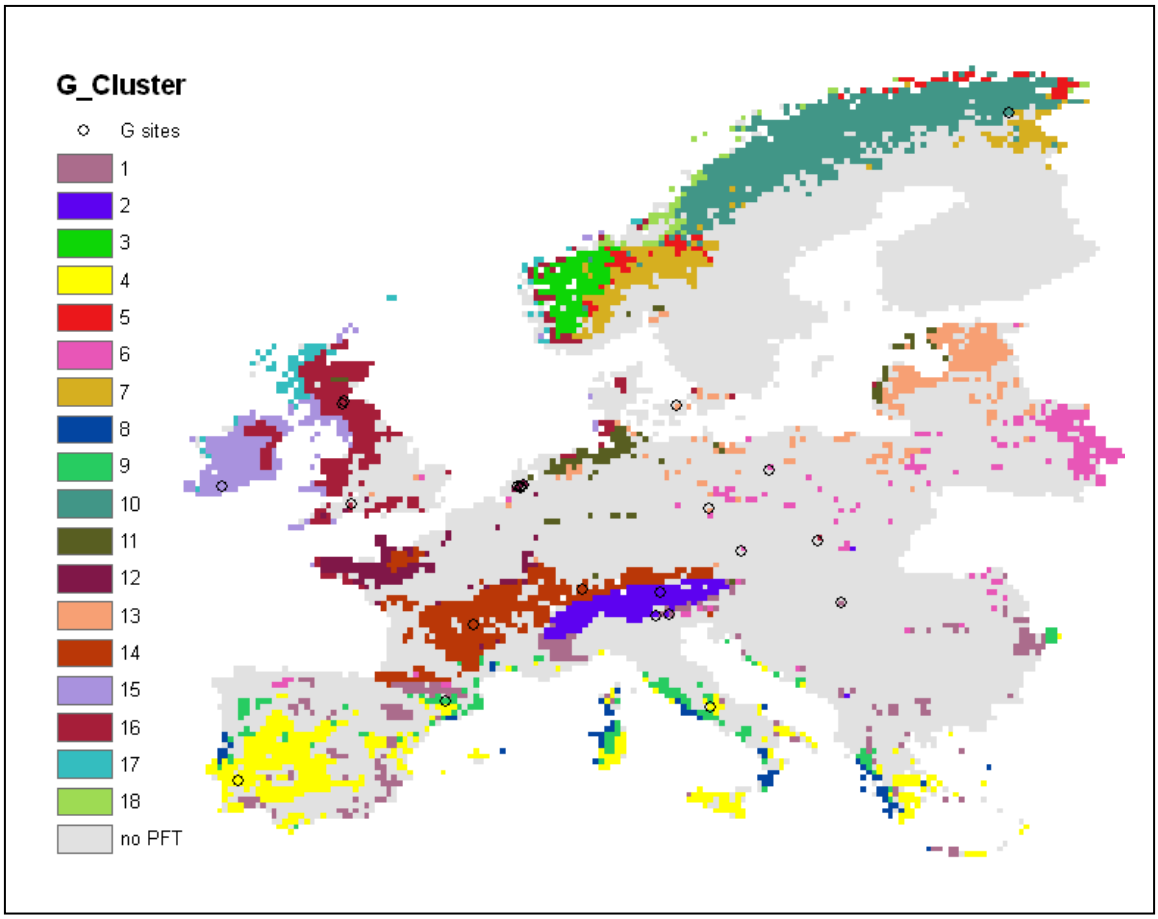


cluster	n°sites	cluster area (Km2)
1	0	2412
2	1	3333
3	0	359
5	0	16871
7	0	5861
8	0	6088
12	0	5569
13	0	23536
15	0	26314
20	0	43667
21	1	17025
23	1	61042
35	0	17405
36	0	14116
37	0	74356
63	0	69119
65	15	812035
66	0	44288
	Total Area	1243398

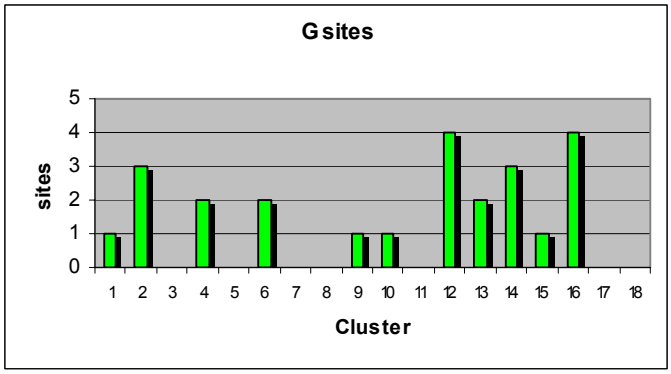


% PFT Area **71.85**

Figura 4.7c. Distribuzione geografica dei cluster e dei siti per ENF

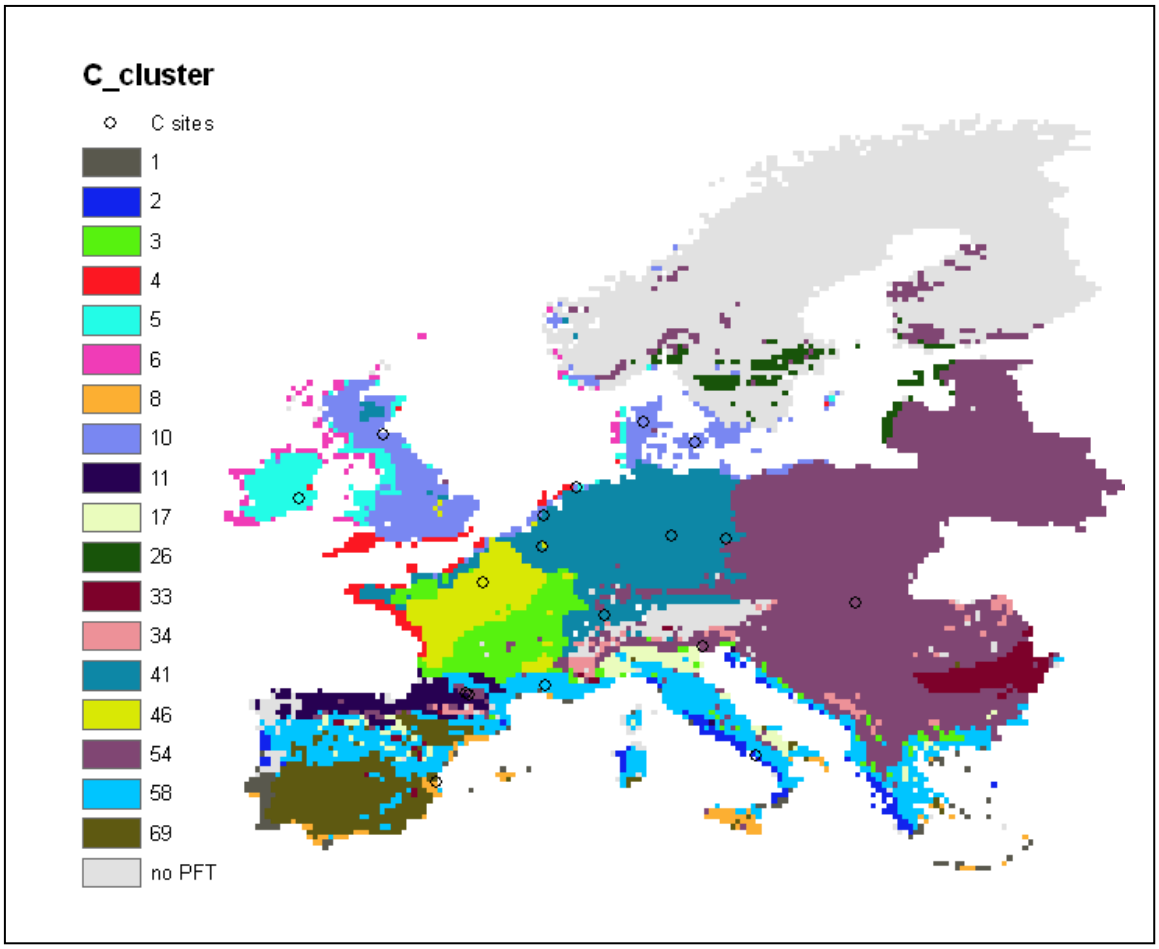


cluster	n°sites	cluster area (Km2)
1	1	9813
2	3	12130
3	0	2611
4	2	14572
5	0	3724
6	2	3669
7	0	4892
8	0	2140
9	1	4370
10	1	27802
11	0	1709
12	4	2902
13	2	4016
14	3	24226
15	1	22290
16	4	12841
17	0	1050
18	0	1173
	Total Area	155932

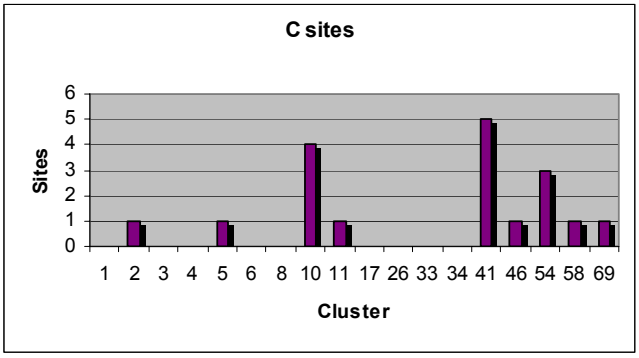


% PFT Area **88.91**

Figura 4.7d. Distribuzione geografica dei cluster e dei siti per G



cluster	n°sites	cluster area (Km2)
1	0	12978
2	1	14066
3	0	106537
4	0	36563
5	1	54903
6	0	13964
8	0	29606
10	4	157666
11	1	31777
17	0	55076
26	0	7057
33	0	78180
34	0	15233
41	5	305706
46	1	140725
54	3	911680
58	1	218425
69	1	162946
	Total Area	2353088



% PFT Area **84.91**

Figura 4.7e. Distribuzione geografica dei cluster e dei siti per C

La mappa geografica relativa alla distribuzione delle foreste latifoglie sempreverdi (EBF, fig. 4.7a) mette in evidenza come i 10 cluster identificati dalle SOM in base ai parametri presi in esame, siano geograficamente abbastanza dispersi nel territorio europeo. Inoltre si evidenzia nettamente come i 10 siti appartenenti alla rete siano distribuiti esclusivamente sui due cluster 14 e 22. Mentre il cluster 22, contenente 7 dei 10 siti, è quello che presenta comunque estensione maggiore, il cluster 14, che pur contiene 3 siti, ha dimensioni piuttosto ridotte (vedi tabella). Ne consegue che l'area rappresentata dai siti EBF ha una bassa percentuale di rappresentatività, solo 57,96%, rispetto all'area totale per questo PFT nel dominio spaziale e relativo ai parametri presi in esame. Tra le tipologie forestali la EBF è quella che presenta una rappresentatività minore.

Il dominio relativo alle foreste latifoglie decidue (DBF, fig. 4.7b), viene invece suddiviso in 19 cluster che occupano aree geografiche meglio determinate. La mappa e la tabella ad essa associata, come già si evidenzia nelle mappe SOM, presenta due cluster di dimensioni notevolmente più grandi degli altri, i cluster 40 e 44, che sono rispettivamente identificati nell'area dell'Europa centrale e nell'area che comprende Spagna e parte dell'est europeo; la metà dei 19 siti totali è distribuita principalmente all'interno di questi due cluster (rispettivamente 7 e 2 siti). La bassa percentuale di area coperta da cluster contenenti siti della rete di monitoraggio, il 58,99%, è determinata dalla presenza di cluster di notevoli dimensioni che non contengono al loro interno alcun sito, in particolare i cluster 33 e 3 che corrispondono a buona parte dell'est europeo e i cluster 4, 6 e 7 dell'estremo nord Europa.

Le foreste di conifere (ENF, fig. 4.7c) presentano una land cover con una distribuzione più omogenea dei 18 cluster. Si evidenzia un grande cluster, il 65, che da solo comprende più del 65% dell'area relativa alla tipologia vegetazionale (individuabile anche al centro della mappa SOM, fig. 4.6c) e che geograficamente occupa un esteso territorio che va dal nord della Spagna fino all'estremo est dell'Europa, attraversando tutta la parte continentale europea e comprendendo anche una buona parte della penisola scandinava e del nord del Regno Unito. In questo cluster sono presenti 15 dei 18 siti totali della rete di monitoraggio, mentre i restanti 3 si trovano divisi in tre cluster distinti. In particolare dalla mappa risulta evidente come due di questi "singoli" siti, IT-SRo e FI-Sod (vedi tabella 2.1a), si trovano rispettivamente all'interno dei cluster 21 e 23, di dimensioni relativamente estese e che geograficamente si dispongono lungo la costa mediterranea il primo, e nel nord della Finlandia il secondo. Questa disposizione dei siti non appartenenti al cluster 65 è determinante per l'aumento dell'area coperta, che nelle conifere raggiunge il 71.85% dell'area totale della PFT. I due siti

sopraindicati sono quindi da considerarsi dei siti "strategici" (singoli siti che occupano cluster di rilevanti dimensioni), che in un eventuale ripensamento della distribuzione della rete non andrebbero comunque spostati fuori dal territorio del cluster. Anche in questo caso si notano cluster di ampie dimensioni, in particolare il 37 e il 63, in cui un eventuale posizionamento di un sito di monitoraggio in foreste di conifere, aumenterebbe notevolmente la percentuale dell'area coperta dai cluster con siti.

Il Land Cover relativo alle praterie diviso dalle mappe SOM in 18 cluster, risulta essere piuttosto frazionato nel territorio europeo (fig. 4.7d). I cluster di maggiori dimensioni si individuano nella penisola scandinava, in centro Europa tra la Francia e la Svizzera, e nel Regno Unito. I 24 siti della rete di monitoraggio si distribuiscono notevolmente tra i cluster; solo 7 cluster su 18 sono privi di siti e sono quelli di minori dimensioni. Questo porta ad un valore molto alto dell'area coperta dai cluster contenenti i siti, che raggiunge quasi il 90% dell'area totale del PFT (88.91%). La tipologia delle praterie è quindi quella che nello studio evidenzia una rappresentatività più alta. Da evidenziare i 4 siti che ricadono nel cluster 12 che è presente in Est Europa in modo estremamente frazionato. Anche in questo caso sono presenti dei siti fondamentali che singolarmente si trovano in cluster di dimensioni notevoli; il sito FI-Kaa nel nord Europa che cade nel cluster 10 e il sito IE-Dri (tab. 2.1a) che in Irlanda cade nel cluster 15 da soli "coprono" più del 32% dell'area dei cluster del dominio per le praterie.

Le cropland occupano il dominio più esteso tra le 5 tipologie vegetazionali studiate (fig. 4.7e). Esso è diviso in 18 cluster nettamente delimitati e non frazionati sul territorio. L'alta omogeneità è dovuta alla notevole diffusione dei terreni coltivati che non tiene conto in questo studio della stagionalità, del management e della estrema variabilità delle singole colture. I cluster sono in generale di grandi dimensioni e i 18 siti sono distribuiti in modo abbastanza omogeneo. I cluster privi di siti sono in generale quelli di minori dimensioni. Fa eccezione il cluster 3, posizionato nel territorio francese, che è privo di siti pur avendo una superficie ragguardevole. Il posizionamento di un sito di monitoraggio su colture nell'area relativa a questo cluster aumenterebbe notevolmente l'area rappresentata, anche se questa è già pari all'84.91% dell'area relativa al PFT. Un sito fondamentale è quello che invece si trova unico nel cluster 58, FR-Avi (tab. 2.1b), nel sud della Francia.

In generale quindi i siti forestali coprono una percentuale minore dell'area relativa ai loro PFT, rispetto alle praterie e alle colture, dove la distribuzione dei siti nel proprio dominio è decisamente maggiore.

Osservando infine i cluster che nelle diverse PFT non contengono siti e confrontando le loro distribuzioni geografiche riportate in figura 4.8, si nota come alcune aree, anche di significative dimensioni, non risultano rappresentate in nessuna delle tipologie prese in esame: esse si distribuiscono soprattutto nel nord Europa, nel sud della Norvegia e nel nord del Regno Unito. A prescindere dalle tipologie vegetazionali dunque, queste sono aree in cui in una nuova disposizione della rete andrebbero posizionati nuovi siti di monitoraggio. Questo risultato concorda con quello discusso nel paragrafo 3.5.1 e riportato in figura 3.6: per queste regioni, entrambe le tecniche applicate evidenziano dunque una distribuzione geografica insufficiente.

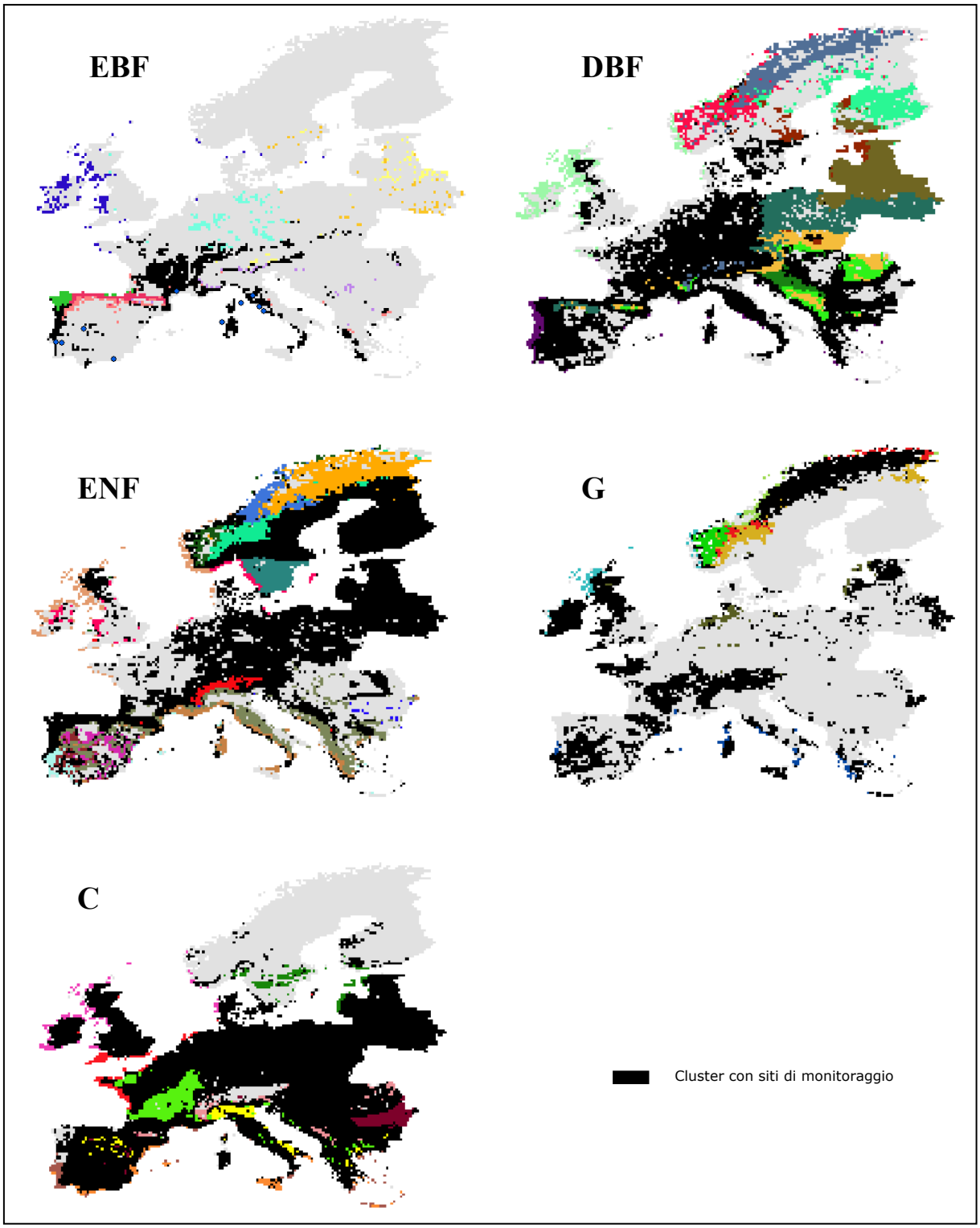


Figura 4.8. Cluster non rappresentati dai siti di monitoraggio

4.4.3 I siti teorici

La creazione dei cluster determinata dall'applicazione del modello SOM non tiene conto della variabilità che si ha all'interno del singolo cluster prodotto. La distanza tra le celle di un singolo cluster è certamente minore di quella che si ha tra un cluster e l'altro, ma per cluster di grandi dimensioni questa può essere rilevante. Nella valutazione della rappresentatività quindi, non si può considerare semplicemente che la distribuzione dei siti deve essere la più ampia possibile. Il peso che ha un sito presente in un cluster molto piccolo non può essere lo stesso di quello che appartiene ad un cluster di più grandi dimensioni. E' quindi chiaro che il numero dei siti che dovrebbero ricadere in un cluster per ottimizzare la rappresentatività dell'insieme dei siti per un particolare PFT, deve essere in qualche modo legato alle dimensioni del cluster, per tener conto anche della variabilità all'interno dello stesso.
Dopo aver determinato il numero dei siti che realmente cadono all'interno dei cluster, si è proceduto quindi a valutare il numero teorico dei siti che dovrebbero appartenere ad un cluster in base alla sua estensione. Per fare questo, per ogni tipologia vegetazionale si è sommata l'area totale dei cluster e la si è divisa per il numero totale dei siti. Si è calcolata in questo modo l'area ottimale per ogni sito con cui poi è stata divisa l'area di ogni cluster. Il risultato è un numero che indica quanti siti (a volte molto meno dell'unità) dovrebbero appartenere ad un determinato cluster in base alla sua dimensione. Prendendo poi i cluster cha hanno un valore di sito teorico (rapporto tra area del cluster e area ottimale per un sito) maggiore di 0.8, viene calcolata di nuovo la percentuale di area coperta dai siti se questi fossero distribuiti secondo il numero dei siti teorici.

Nella figura 4.9 sono mostrati i grafici relativi alla distribuzione dei siti reali e teorici per ogni PFT. Confrontando le distribuzioni si nota come per le tipologie che hanno una bassa percentuale di rappresentatività, la distribuzione teorica si allontana di molto da quella reale, che generalmente è molto concentrata su un minor numero di cluster: questo è particolarmente evidente nelle EBF dove i siti reali sono concentrati in soli due cluster mentre la distribuzione teorica prevede siti distribuiti in altri due cluster (1 e 2) mentre quasi non ne prevede (valore 0.28) per il cluster 14 dove invece ci sono realmente ben 3 siti. Per le tipologie che hanno invece una percentuale di area coperta dai siti già abbastanza elevata, la distribuzione teorica si avvicina a quella reale. Caso particolare è presentato per le praterie, dove i siti reali sono più dei cluster individuati dalle SOM (24 siti per 18 cluster) e quindi l'area ottimale per ogni sito risulta minore; l'elevata presenza però di cluster di piccole dimensioni porta ad una

distribuzione dei siti teorici tale che in molti cluster il rapporto area del cluster/area ottimale non arriva al valore di 0.8 individuato come soglia per attribuire al cluster stesso un sito.

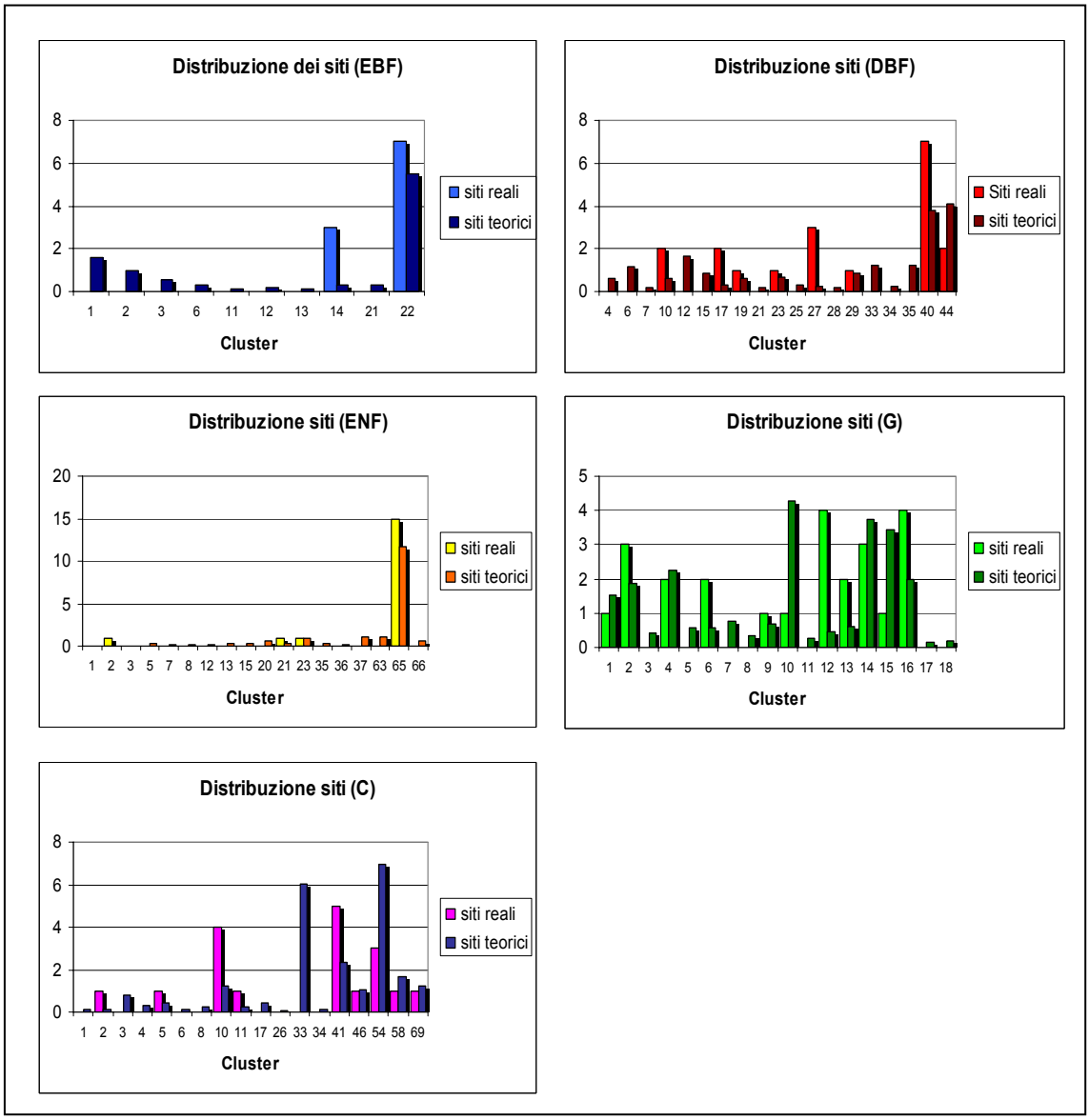


Figura 4.9. Distribuzione dei siti reali e teorici all'interno dei cluster

Calcolando la percentuale di area coperta da ogni singolo PFT in base alla distribuzione dei siti teorici, assegnando un sito ad un cluster quando il valore del rapporto tra area del cluster e area ottimale è maggiore di 0.8, si ottengono i risultati riportati in tabella 4.2:

	siti reali	siti teorici	area teorica /sito (KM2)	% PFT coperta con siti reali	% PFT coperta con siti teorici	diff. %
EBF	10	7	3022	57.96	80.87	22.91
DBF	19	13	34581	58.99	77.71	18.72
ENF	18	14	69078	71.85	81.76	9.91
G	24	17	6497	88.91	79.31	-9.6
C	18	20	130727	84.91	88.47	3.56

Tabella 4.2. Confronto tra % di copertura per siti reali e siti teorici

La ridistribuzione dei siti teorici porta un aumento notevole della percentuale di PFT coperta dai cluster per le EBF. Con 7 soli siti (a fronte dei 10 reali) distribuiti però nei cluster 1, 2 e 22 (fig. 4.9), l'area coperta dai cluster arriva al valore di 80.87% aumentando più di 22 punti percentuale. Anche per le DBF il numero dei siti teorici è minore di quello dei reali (13 invece di 19) ma l'area coperta sale al 77.71% dell'area totale della PFT grazie soprattutto alla distribuzione di siti all'interno di cluster di rilevanti dimensioni come il 6, il 12, il 33 e il 35 (questi ultimi due già notati risultanti determinanti nell'analisi delle mappe geografiche del paragrafo precedente). Le conifere (ENF) sono la tipologia forestale che, avendo già una percentuale di rappresentatività del 71.85%, aumenta solo di quasi 10 punti percentuale la sua rappresentatività teorica con un totale di 14 siti teorici. L'aumento è dovuto all'assegnazione di un sito ai cluster 37 e 63, geograficamente posizionati entrambi nella penisola scandinava che viene individuata quindi come una regione dove sarebbe importante aggiungere siti di monitoraggio su foreste di conifere (par.4.4.2).

Le praterie sono le uniche che, come già descritto in precedenza, per la dimensione dei cluster in cui è diviso il territorio a cui appartengono, mostrano una distribuzione teorica troppo frammentata e la percentuale di copertura teorica per loro risulta essere più bassa di quella reale. In questo caso infatti i 24 siti reali risultano essere ridondanti per il numero dei cluster individuati dalle SOM e questo comporta una minore estensione dell'area teorica prevista per ogni sito. Un' ottimizzazione della rete prevede quindi una diminuzione del numero dei siti da 24 a 17.

Le colture risultano invece avere la distribuzione reale più vicina a quella teorica aumentando solo di 3 punti e mezzo sulla percentuale di rappresentatività reale. Questa tipologia vegetazionale è l'unica in cui il numero dei siti teorici aumenta rispetto a quello dei reali da 18 a 20. Da notare che nella distribuzione teorica dei siti, viene assegnato un sito al cluster 3 (fig. 4.9) che già nel paragrafo precedente era stato identificato come un cluster determinante per l'aumento della percentuale di rappresentatività.

In definitiva il calcolo dei siti teorici mostra in generale che una diversa distribuzione dei siti in base alla dimensione dei cluster potrebbe permettere la riduzione del numero dei siti stessi all'interno della rete di monitoraggio.

4.4.4 Gli "hotspots"

Come è stato già sottolineato nel precedente paragrafo esiste una certa variabilità all'interno di ogni cluster; le celle appartenenti ad un singolo cluster sono vicine tra loro nello spazio a 21-D ma non sono uguali. E' possibile quindi individuare per ogni cluster il suo centro, inteso come la cella che meglio rappresenta le caratteristiche dell'intero cluster stesso, combinando due criteri: la quantizzazione dell'errore e la superficie del cluster. Una volta individuato il centro (paragrafo 4.3.2) e la distanza euclidea di ogni pixel da questo, la distanza stessa viene usata come "misura di qualità": viene creato quindi un indice di distanza associato ad ogni pixel, che varia tra 0 e 1 dove 1 indica un "hotspot", cioè un pixel con la massima vicinanza al centro del cluster, che quindi rappresenta una buona area per posizionare un sito.
Il risultato che si ottiene per ogni PFT è una mappa dove ogni pixel appartiene ad una scala relativa a questa "misura di qualità" (fig. 4.10). I pixel con bassi valori sono quelli in cui è meno indicato posizionare un sito, mentre i pixel con alti valori contraddistinti dal rosso sono quelli più indicati per il posizionamento di un sito. I siti che sono posti in questi "hotspots" sono appunto quelli che risultano essere più vicini al centro del cluster e quindi quelli più rappresentativi.

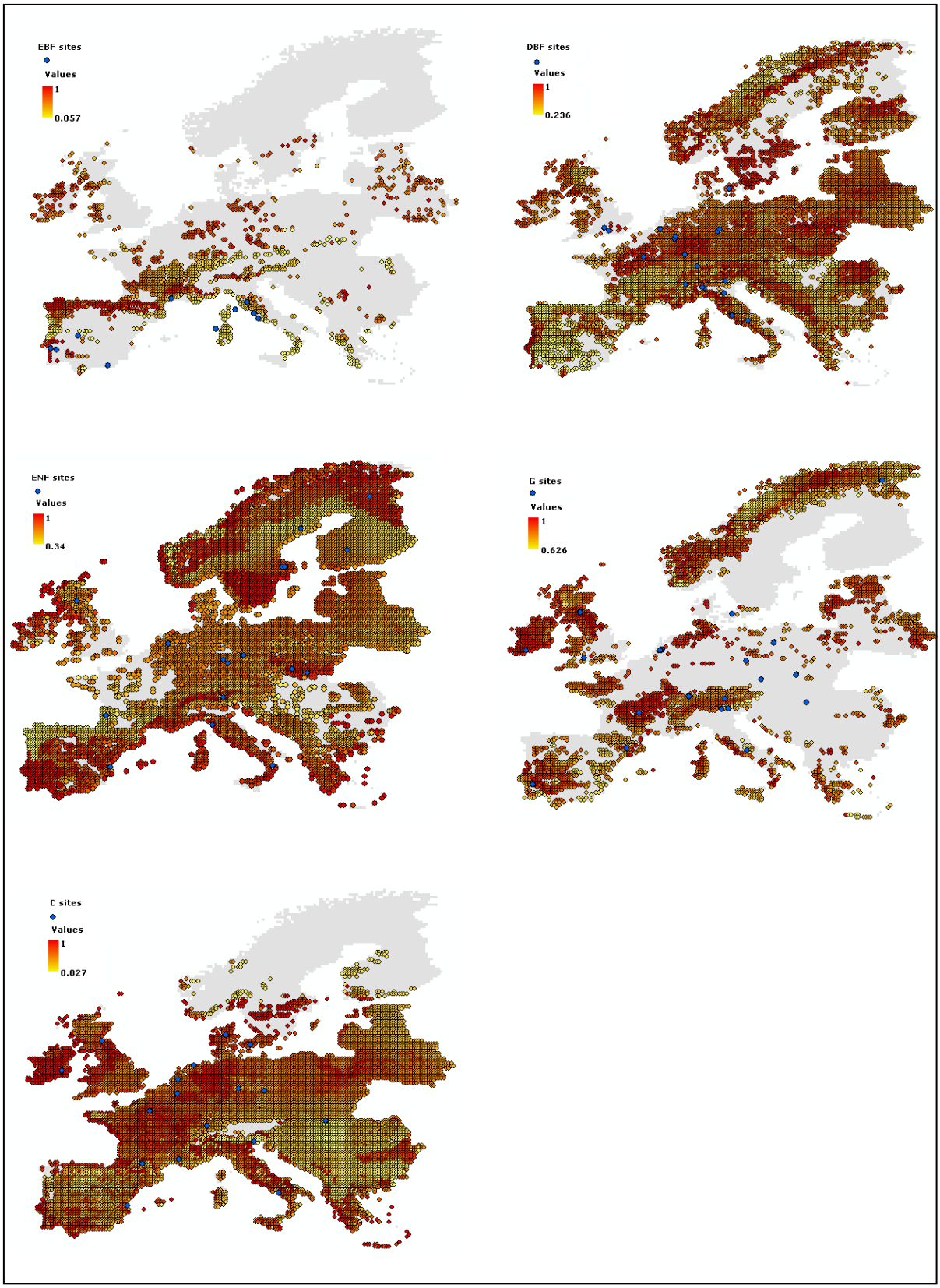


Figura 4.10. Mappa degli "hotspots" indicanti i centri dei cluster

Posizionando sulle mappe i siti dell'attuale rete di monitoraggio, per ogni PFT si possono individuare quali siti sono più vicini ai centri dei cluster che li contengono e quindi sono più rappresentativi all'interno di un cluster. I risultati vengono anche riassunti nelle tabelle di figura 4.11, dove per ogni sito viene indicato il valore della sua distanza normalizzata dal centro del cluster a cui appartiene: valori prossimi ad uno indicano che il sito rappresenta bene il cluster di appartenenza, mentre valori prossimi allo zero indicano che il sito diverge molto dalle condizioni trovate al centro del cluster. In arancio vengono evidenziati i siti che hanno un valore maggiore di 0.9.

Osservando le mappe riportate in figura 4.9 e i valori in figura 4.10 è evidente come tra le diverse tipologie vegetazionali le foreste di conifere siano quelle che presentano un numero maggiore di siti della rete di monitoraggio, vicini ai centri dei cluster determinati con le reti SOM (siti posti in aree più rosse e con un valore della distanza normalizzata maggiore di 0.9): questo comunque non determina una particolare alta percentuale di rappresentatività dei siti ENF in generale (come infatti viene descritto nel paragrafo 4.4.2) perché più siti sono vicini nel dominio a 21-D al centro dello stesso cluster (in questo caso il cluster 65). Allo stesso modo si nota come per le praterie che pur hanno una percentuale di copertura dei cluster che contengono siti molto alta (vicina al 90%), solo 3 siti su 24 sono molto rappresentativi del centro del cluster. Le EBF presentano solo un sito con un valore più alto di 0.9; il sito PT-Mi2 appartenente al cluster 14 è l'unico che si presenta vicino al centro del cluster di appartenenza, mentre tutti i 7 siti appartenenti al cluster 22 sono comunque lontani dal suo centro nelle 21-D.

Le mappe degli hotspots vanno quindi interpretate combinandole con quelle già date per la distribuzione geografica dei cluster, e là dove un cluster sia risultato determinante per l'aumento della percentuale di copertura ma privo di siti, indicano dove è posto il centro del cluster stesso e quindi la migliore posizione di un nuovo sito. Per le EBF, per esempio, dove la distribuzione geografica dei siti e quella dei siti teorici ha evidenziato una mancanza di siti nei cluster 1 2 e 3, la mappa degli hotspots individua nel nord della Spagna e dell'Irlanda la migliore posizione per un nuovo sito EBF.

EBF	Value
ES-LMa	0.66
ES-LJu	0.14
FR-Pue	0.76
IT-Cpz	0.65
IT-Tol	0.67
IT-Pia	0.61
IT-Noe	0.56
IT-Lec	0.76
PT-Mi1	0.92
PT-Esp	0.85

DBF	Value
BE-Vie	0.75
BE-Bra	0.77
BE-Jal	0.84
CH-Lae	0.48
DE-Hai	0.81
DE-Meh	0.77
DK-Sor	0.53
FR-Hes	0.89
FR-Fon	0.80
IT-Col	0.37
IT-Ro1	0.99
IT-Non	0.91
IT-PT1	0.71
IT-Lav	0.90
IT-Ro2	0.99
IT-LMa	0.53
IT-Vig	0.75
UK-PL3	0.67
UK-Ham	0.76

ENF	Value
CZ-BK1	0.98
DE-Bay	0.97
DE-Tha	0.98
DE-Wet	0.98
ES-ES1	0.79
FI-Hyy	0.96
FI-Sod	1.00
FR-LBr	0.91
IT-Ren	0.41
IT-SRo	0.93
IT-Bon	0.87
NL-Loo	0.96
SE-Nor	0.98
SE-Fla	0.92
SE-Sk1	0.98
SE-Sk2	0.98
Sk-Tat	0.97
UK-Gri	0.96

G	Value
AT-Neu	0.93
CH-Oe1	0.70
CZ-BK2	0.19
CZ-wet	0.54
DE-Gri	0.49
DK-Lva	0.16
ES-VDA	0.82
FI-Kaa	0.97
FR-Lq1	0.87
FR-Lq2	0.87
HU-Bug	0.40
IE-Dri	0.75
IT-Amp	0.39
IT-MBo	0.74
IT-Mal	0.92
NL-Ca1	0.68
NL-Hor	0.57
NL-Ca2	0.68
NL-Haa	0.76
PL-wet	0.87
PT-Mi2	0.82
UK-Tad	0.76
UK-EBu	0.44
UK-AMo	0.44

C	Value
BE-Lon	0.98
CH-Oe2	0.96
DE-Geb	0.93
DE-Kli	0.96
DK-Ris	0.77
DK-Fou	0.85
ES-ES2	0.96
FR-Gri	0.94
FR-Avi	0.75
FR-Aur	0.88
FR-Lam	0.40
HU-Mat	0.96
IE-Ca1	0.94
IT-BCi	0.89
IT-Neg	0.95
NL-Lan	0.97
NL-Lut	0.84
UK-ESa	0.87

Figura 4.11. Distanza normalizzata dei siti della rete dai centri dei cluster

Analizzando invece la mappa degli hotspots per le DBF si può individuare l'area più rappresentativa per posizionare un sito nell'Europa dell'est, là dove i cluster 33 e 35 della distribuzione geografica ne risultano privi.

Per le conifere le mappe degli hotspots mettono in evidenza come i centri dei rilevanti cluster privi di siti, 3, 63 e 20 siano posizionati nella parte dell'estremo nord Europa (Norvegia in particolare), mentre per il cluster 66, frammentato geograficamente tra Spagna, Italia e Grecia

non risulta evidente dove sia posizionato il centro del cluster. Inoltre i due siti che nella distribuzione geografica risultavano di importanza rilevante perché rappresentativi di cluster molto grandi, IT-SRo e FI-Sod (par. 4.4.2), risultano anche molto ben posizionati rispetto al centro del proprio cluster (valore di 0.93 e 1.00 rispettivamente, nella figura 4.11).
All'interno del dominio delle praterie invece, diversa è la situazione dei due siti che sono risultati importanti per la loro appartenenza a cluster estesi: FI-Kaa si trova ben posizionato vicino al centro del suo cluster, mentre IE-Dri ha una distanza normalizzata pari a 0.75, e quindi dovrebbe essere spostato maggiormente verso ovest, dove si trova il centro del cluster 15 a cui il sito appartiene.
Stesse considerazioni si possono fare per il sito FR-Avi che risultava essere, tra i siti appartenenti ai coltivi, quello fondamentale per il cluster 58: anche ad esso risulta associata una distanza normalizzata pari a 0.75, il centro del suo cluster però è meno evidente e sembra posizionato più sulla parte centrale dell'Italia.

In generale quindi la mappe degli hotspots aggiungono informazioni sul posizionamento dei centri dei cluster soprattutto quando questi sono ben delimitati e poco frammentati, mentre danno indicazioni meno chiare là dove i cluster sono distribuiti in modo frammentato sul territorio europeo. In questo caso l'analisi dovrebbe procedere all'interno di ogni PFT considerando e visualizzando separatamente ogni singolo cluster, in modo da fornire risposte anche visivamente più chiare sul centro di ciascun cluster.

4.5 Conclusioni

L'applicazione del modello SOM al dominio a 21-D costituito dalle variabili climatiche e ecofisiologiche ha permesso per ogni tipologia vegetazionale studiata, di suddividere il territorio europeo in cluster di condizioni simili e per ognuno dei Plant Functional Type è stato possibile individuare quali variabili hanno pesato di più sulla costituzione di questi cluster (par. 4.4.1). Riportando i cluster su scala geografica è stata individuata la posizione dei siti della rete di monitoraggio all'interno degli stessi ed è stata fatta un'analisi della loro rappresentatività in base alla loro appartenenza ai singoli cluster (par. 4.4.2). Ne è risultato che per quanto riguarda le tipologie forestali considerate, i siti di latifoglie sempreverdi (EBF) sono quelle hanno una rappresentatività minore, dovuta sia alla presenza di un numero insufficiente di siti, sia al fatto che quelli presenti siano concentrati su due soli cluster; i siti di

conifere (ENF) sono invece i siti forestali con una maggiore rappresentatività, dovuta sia alla presenza dei siti in un cluster di notevoli dimensioni posizionato al centro dell'Europa, sia alla presenza di siti "strategici", siti cioè che singolarmente si trovano all'interno di cluster di dimensioni rilevanti. Per i siti non forestali, le praterie e le colture, è stata individuata una rappresentatività in generale molto più alta, essendo notevolmente più distribuiti a coprire i cluster individuati dal modello SOM. Dall'analisi della distribuzione geografica è quindi possibile anche identificare quali cluster e quindi quali regioni del territorio europeo sono attualmente prive di siti e quali tra questi risultano importanti per un aumento della rappresentatività. In particolare per le EBF sono state individuate come sede di nuovi siti in una ipotetica ridistribuzione della rete, il nord della Spagna, l'Irlanda del Nord e la Germania; per le DBF aree prive di siti si sono rivelate l'Est europeo e la Norvegia; nuovi siti di conifere potrebbero essere posizionati nel sud della Svezia, in Norvegia e nella regione mediterranea; infine almeno un nuovo sito di colture potrebbe essere individuato nella parte centrale della Francia.

Dalla distribuzione dei siti e dalle aree dei singoli cluster è stato possibile quindi ipotizzare una nuova distribuzione di siti teorici (par.4.4.3) che in generale ha delineato per ogni PFT una nuova rete costituita da un minor numero di siti distribuiti però maggiormente all'interno dei cluster; fanno eccezione a questo le colture per le quali il numero di siti teorici è aumentato di due unità rispetto ai siti reali. Per tutte le tipologie vegetazionali la percentuale di area coperta dai cluster contenenti i siti teorici è aumentata rispetto a quella calcolata con i siti reali, eccezione fatta per le praterie dove la piccola dimensione dei singoli cluster determina una riduzione della percentuale dell'area coperta con i siti teorici.

La variabilità all'interno dei singoli cluster è stata analizzata individuando nello spazio 21-D di un determinato PFT il centro di ogni cluster (par.4.4.4). Sono state prodotte mappe cosiddette di "hotspots" dove si evidenziano le zone che essendo più "vicine" al centro di ogni cluster sono quindi le più rappresentative dello stesso e quindi le più idonee ad ospitare un sito di monitoraggio in una ipotetica nuova rete. Confrontando i siti attualmente presenti con questi hotspots, risulta evidente che per alcuni PFT molti siti si trovano in buone posizioni, vicini a centri di cluster (le ENF hanno il numero di "buoni siti" maggiore) mentre altre tipologie vegetazionali, come le EBF, hanno i siti lontani dai centri. Questo non influisce comunque direttamente sulla rappresentatività della rete attuale, in quanto diversi siti possono essere vicini al centro di uno stesso cluster, come infatti avviene per le ENF, e quindi risultare in qualche modo ridondanti.

In generale quindi l'applicazione del modello SOM ha permesso di investigare la rappresentatività all'interno di un dominio comprendente 21 dimensioni, di valutare la posizione dell'attuale rete di monitoraggio in questo dominio e di dare indicazioni su possibili nuove distribuzioni dei siti.

Capitolo 5

Conclusioni finali e sviluppi

Le tecniche di analisi utilizzate in questo studio per valutare la rappresentatività della rete dei siti di monitoraggio afferenti al progetto CarboeuropeIP, hanno permesso di raggiungere alcuni risultati importanti in questo primo livello investigativo.
Le due analisi applicate hanno portato entrambe alla individuazione di aree omogenee all'interno del territorio europeo, entro le quali è stato possibile valutare distribuzione e rappresentatività delle torri di monitoraggio.

L'analisi GIS (capitolo 3) ha consentito la creazione di 57 Homogeneous Territory Unit (HTU) più o meno estese o frammentate, su ognuna delle quali sono stati confrontati i dati delle stazioni presenti sul proprio territorio; il numero delle HTU individuate deriva dalla classificazione delle variabili spazializzate del dataset e dipende quindi dalla risoluzione spaziale dei dati stessi. Analizzando la distribuzione spaziale dei siti, essi ricadono in 18 HTU che coprono il 74.78% dell'area totale, indice di una buona rappresentatività geografica della rete. Sono state individuate tuttavia aree, anche di estese dimensioni, che risultano prive di siti, quali la penisola scandinava e il nord della Francia.
Analizzando inoltre l'Indice di Coincidenza che viene calcolato confrontando i dati delle HTU con quelli puntuali provenienti dalle torri, si sono raggiunti due risultati. In primo luogo nel dominio considerato è stato possibile valutare la coincidenza dei siti con le HTU di appartenenza: l'83.15% dei siti ha con la propria HTU un IC maggiore di 0.8., e questo dimostra una buona correlazione tra i parametri misurati nei siti di monitoraggio e quelli derivati dai dati spazializzati. In secondo luogo, analizzando l'Indice di Coincidenza dei siti a prescindere dalla HTU di appartenenza, si è evidenziata la presenza di una rappresentatività non geografica, ma "ecofisiologica" (par. 3.5.3); alcuni siti infatti si dimostrano rappresentativi di HTU in cui essi non sono geograficamente posizionati. Questo conferma che il concetto di rappresentatività non è necessariamente legato appunto alla posizione geografica del sito preso in esame, ma strettamente definito dal dominio che viene scelto per la sua valutazione.

L'analisi GIS ha posto in rilievo la presenza di alcuni limiti applicativi (par. 3.6) dovuti alla risoluzione dei dati spazializzati su scala continentale, che non risulta sufficiente per un confronto con i dati puntuali, soprattutto in aree più frammentate del territorio europeo.

E' stata quindi condotta la seconda analisi utilizzando il modello che si avvale dell'uso delle SOM (capitolo 4). L'applicazione del modello ha permesso di estendere il dominio su cui valutare la rappresentatività e da un insieme di 5 variabili si è passati a considerare uno spazio a 21 dimensioni. Inoltre l'analisi è stata condotta considerando separatamente 5 Plant Functional Type (tre di tipo forestale, praterie, coltivi), e questo ha portato ad una valutazione non vincolata strettamente alla risoluzione dei dati spazializzati.
L'analisi ha permesso la suddivisione del territorio occupato da ogni PFT, in cluster omogenei nello spazio a 21-D. All'interno di questi cluster che occupano aree continue di matrici bidimensionali, è stata analizzata la distribuzione dei siti.
Attraverso questa fase dello studio è stato possibile quindi valutare, per ognuna delle tipologie vegetazionali considerate, il grado di rappresentatività dei siti che le appartengono, potendo constatare come le tipologie forestali siano in generale meno rappresentate delle praterie e delle colture (par. 4.4.2). E' stato quindi possibile anche verificare quali aree geograficamente risultano in ogni caso meno rappresentate, a prescindere dai PFT di appartenenza. Queste aree, collocate geograficamente nel nord Europa (sud della Norvegia e nord del Regno Unito) coincidono con le maggiori HTU che nella precedente analisi GIS venivano indicate come prive di siti e quindi con scarsa rappresentatività geografica. L'individuazione di queste aree che emerge da ciascuna delle tecniche applicate, rappresenta un primo criterio da considerare in una eventuale disposizione di nuovi siti all'interno della rete.
Dalla relazione dei siti e dei cluster è stato quindi possibile ipotizzare per ogni tipologia vegetazionale una nuova distribuzione dei siti, che possono essere in generale anche ridotti di numero. La definizione di siti "teorici" ha permesso, nel caso delle foreste di latifoglie sempreverdi, di aumentare la rappresentatività geografica anche di più di 20 punti percentuale.
Un ulteriore approfondimento dell'analisi ha permesso infine di individuare nel territorio europeo posizioni "hotspots", aree in ogni cluster più rappresentative delle sue caratteristiche, dove l'eventuale posizionamento di un sito aumenterebbe notevolmente la rappresentatività dello stesso e dell'intera rete. Questa informazione può essere aggiunta alla precedente individuazione di aree lasciate in qualche modo "non rappresentate" dai siti.

In generale si può affermare che per lo studio della rappresentatività su scala continentale l'analisi SOM si è dimostrata in grado di definire un complesso dominio, costituito da un numero elevato di variabili, e ha permesso un dettagliato confronto tra i dati spazializzati e quelli puntuali dei siti. La sua capacità nel poter rappresentare in modo relativamente semplice uno spazio a n-dimensioni, ne fa uno strumento potente nell'analisi multivariata.

Sulla base dei risultati ottenuti in questo lavoro, è stato possibile dare indicazioni per migliorare la rete, per essere in grado in futuro di tracciare un suo ridisegnamento, sia aumentando il numero dei siti, posizionandoli nelle aree più rappresentative, sia riposizionando eventualmente quelli già esistenti in modo tale che l'intera rete possa arrivare a rappresentare un'area pari almeno al 90% di quella appartenente all'intero territorio europeo. Una rete maggiormente rappresentativa permetterà analisi di valutazione sugli ecosistemi presenti in Europa, tendenti a monitorare e quantificare gli scambi di carbonio tra superficie ed atmosfera, e consentirà di prevedere azioni sulla variabilità e sulla gestione antropica degli ecosistemi stessi.

5.2 Criticità

Come già evidenziato e discusso nel capitolo 3 e nella prima parte di questo capitolo, la risoluzione dei dati spazializzati utilizzati nelle analisi, unita alla estrema frammentarietà del paesaggio europeo, pone dei limiti per le analisi di classificazione su scala continentale.
In questo studio inoltre non sono state considerate alcune variabili che influenzano a volte in modo deciso le caratteristiche di un determinato territorio e la tipologia vegetazionale presente. Non sono stati presi in considerazione per esempio dati relativi al suolo, come la temperatura o il contenuto idrico, e non sono state utilizzate, tra i dati spazializzati, mappe di uso del suolo. Nei dati raccolti dalle torri eddy infatti queste informazioni sono spesso mancanti o non complete e ciò impedisce il loro uso nella definizione del dominio considerato. Altre variabili che potrebbero risultare fondamentali soprattutto per alcune tipologie vegetazionali, sono il management dei siti di monitoraggio e l'impatto antropico, che determinano una notevole variabilità non solo di tipo temporale ma anche spaziale, tra un sito e l'altro all'interno di un singolo Plant Functional Type. Questa variabilità è particolarmente determinante nei siti di monitoraggio relativi ai coltivi, che hanno una stagionalità e un management estremamente diversificati tra loro. L'impossibilità di

quantificare variabili come il management e l'impatto antropico ne ha impedito di fatto l'uso nelle analisi e ha comportato una omogeneità e una maggiore estensione dei cluster relativi soprattutto ai coltivi e alle praterie, con una possibile sovrastima della rappresentatività dei siti situati in essi.

5.1 Prospettive

Lo studio sulla rappresentatività presentato, dimostra come questa sia altamente legata al concetto di dominio entro il quale essa si definisce. Le due analisi costituiscono un primo approccio nella valutazione della rete dei siti di monitoraggio facenti parte del progetto CarbouropeIP e pone le basi per approfondimenti e implementazioni future.

In primo luogo il dominio può essere ampliato e modificato, in particolare per i dati spazializzati, dove le variabili che definiscono la componente ecofisiologica possono essere desunte da altre fonti per verificare la validità dei risultati ottenuti. Molti modelli oggi in via di implementazione forniscono dati di GPP e NEE su scala continentale (Vetter, 2008): LPJ e Orchidee sono solo due dei molti modelli oggi a disposizione che potrebbero essere utilizzati nella definizione di un nuovo dominio.

Inoltre questo lavoro prende in considerazione solo uno degli innumerevoli domini possibili ed ha dato risultati di variabilità relativi ad esso, ma è applicabile a molti altri domini e la rete può essere valutata su altri parametri che potrebbero dare risultati analoghi o diversificati rispetto a quelli ottenuti in questo studio.

Un'ulteriore implementazione dell'analisi potrebbe anche riguardare l'approfondimento dello studio sulla eterogeneità all'interno dei cluster che risultano dall'applicazione delle SOM. Cluster più eterogenei dovrebbero essere coperti da un numero maggiore di siti; la relazione tra eterogeneità dei cluster e numero di siti ad essi relazionati, costituisce un ulteriore criterio da considerare nel riposizionamento della rete.

Bibliografia

Armenteras D., Gast F., Villareal H., (2003) Andean forest fragmentation and the representativeness of protected natural areas in the eastern Andes, Colombia, *Biological Conservation*, 113, 245-256.

Aubinet M., Grelle A., Ibrom A., Rannik U., Moncrieff J., Foken T., Kowalski A., et al., (2000) Estimates of the Annual Net Carbon and Water Exchange of Forest: The EUROFLUX Methodology, *Advances in Ecological Research*, 30, 114–173.

Awimbo J.A., Norton D.A., (1996) An evaluation of representativeness for nature conservation, Hokitika Ecological District, New Zealand, *Biological Conservation*, 75, 177-186.

Biallo G., (2005) Introduzione ai Sistemi Informativi Geografici, 3a ed., Mondogis.

Baldocchi D., Falge E., Gu L., Olson R., Hollinger D., Running S., et al., (2001) FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities, *Bull. Am. Meteorol. Soc.*, 82, 2415–2434.

CarboeuropeIP Project – www.carboeurope.org

DeMers M.N., (2005) Fundamentals of Geographic Information System, 3th Edition, John Wiley & Sons. Ed.

Foken et al., (2003) Post_field data quality control, *Handbook of Micrometeorology*, Ed. Lee et al., 181-208.

Han J., Kamber M., (2001) Data mining: concepts and techniques, Morgan Kaufmann, San Francisco

Hargrove W.W., Hoffman F.M., Law B.E., (2003) New Analysis Reveals Representativeness of the Ameriflux Network, *EOS, Transactions, American Geophysical Union*, 84, n°48, 529-544.

Gomarasca M.A., (1997) Introduzione a telerilevamenteo e GIS per la gestione delle risorse agricole e ambientali, Associazione Italiana di Telerilevamento Ed.

Janssens I.A., Lankreijer H., Matteucci G., (2001) Productivity overshadows temperature in determining soil and ecosystem respiration across European forests, *Global Change Biology*, 7, 269-278.

Jenks G.F., Caspall F.C., (1971) Error on choroplethic maps: definition, measurement, reduction, *Annals of American Geographers*, 61, 217-44.

Jongman R.H.G., ter Braak C.J.F., van Tongeren O.F.R., (1995) Data analysis in community and landscape ecology, Cambridge University Press, Cambridge.

Kaski S., (1997) Data exploration using Self-Organizing Maps, PhD Thesis, Helsinki, University of Technology, Espoo, Finland.

Knapp E., (1978) Landsat and ancillary data inputs to an automated Geographic Information System, Report CSC/TR-78/6019, Computer Science Corp., Silver Spring, USA.

Kohonen T., (1981) Construction of Similarity Diagrams for Phonemes by a Self-Organization Algorithm, Report TKK-F-A463, Helsinki, University of Technology, Espoo, Finland.

Kohonen T., (2001) Self-Organizing Maps, 3th edition, Springer.

Lek S., Giraudel J., Guégan J.F., (2000) Neuronal Network: Algorithms and Architectures for Ecologists, in Artificial Neuronal Networks, Springer.

Luyssaert S., Janssens I.A., Sulkava M., Papale D., Dolman A.J., et al (2007) Photosynthesis drives anomalies in net carbon-exchange of pine forests at different latitudes, *Global Change Biology*, 13, 2110-2127.

Luyssaert S., Sulkava M., Raitio H., Hollmen J., (2004) Evaluation of forest nutrition based on large-scale foliar surveys: are nutrition profiles the way of the future?, *J. Environ. Monit*., 6, 160-167.

O'Connor K.F., Overmars F.B., Ralston M.M., (1990) Land evaluation for nature conservation. A scientific review compiled for application in New Zealand. *Conserv. Sci. Publs*.,3, Department of Conservation, Wellington.

Oja M., Kaski S., Kohonen T., (2003) Bibliography of self-organizing map (SOM) papers: 1998-2001 addendum. *Neural Computing Surveys*, 3, 1-156.

Papale D, Valentini R., (2003) A new assessment of European forests carbon balance exchanges by eddy fluxes and artificial neural network spatialization, *Global Change Biology*, 9, 525-535.

Papale D., Reichstein M., Aubinet M., Canfora E., Bernhofer C., Kutsch W., et al. (2006) Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, *Biogeosciences*, 3, 571-583.

Pyle D., (1999) Data preparation for data mining, Morgan Kaufmann Publishers.

Reichstein M., Tenhunen J., Ourcival J.M. et al., (2002) Ecosystem respiration in two Mediterranean evergreen holm oak forest: drought effects and decomposition dynamics, *Functional Ecology*: 16, 27-39.

Reichstein M., Falge E., Baldocchi D., Papale D., Aubinet M., Berbigier P., et al. (2005) On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biology* , 11, 1424-1439.

Taramelli A., Santini M., Sorichetta A., Algorithm to calculate cell area on a latitude-longitude (geographic) regular grid. Submitted to International Journal of GIS

Ultsch A., Siemon H.P., (1990) in *Proceedings of the international neural network conference* (INNC '90), ed. I.N.N.C., Kluwer Dordrecht, 305-308.

Valentini R., et al., (2003) Fluxes of carbon, water and energy of European forests, *Ecological Studies*, Springer, vol 163.

Vellido A., Lisboa P.J.G., Meehan K., (1999) Segmentation of the on-line shopping market using neural networks, *Expert System with Applications*, 17, 303-314.

Vesanto J., Sulkava M., (2002) Distance matrix based clustering of the Self-Organizing Map, *Lect. Notes Sci.*, 2415, 951-956.

Vistocco D., (2002) Mappe di Kohonen, in: *Reti Neuronali e Metodi Statistici*, Ed. Angeli: 175-191.

Vetter M., Churkina G., Jung M., Reichstein M., Zaehle S., Bondeau A., et al., (2008) Analyzing the causes and spatial pattern of the European 2003 carbon flux anomaly using seven models, *Biogeosciences*, 5, 561-583.

Bibliografia non citata

Baldocchi D., (2003) Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, *Global Change Biology*, 9, 479-492.

Burrough P.A., McDonnell R.A., (1998) Principles of Geographical Information Systems, New York: Oxford University Press.

Chrisman N., (1987) The accuracy of map overlays: a reassessment, *Landscape and urban planning*, 14, 427-439.

DeMers M.N., (2002) GIS modeling in raster. New York: John Wiley & Sons.

Kiviluoto K., (1996) Topology preservation in self-organizing maps. In: *Proceedings of the International Conference on Neural Networks* (ICNN'96), Vol. 1 (ed. CINN), Piscataway, New Jersey 294-299.

Lillesand T.M., Kiefer R.W., (1995) Remote sensing and image interpretation, New York: John Wiley & Sons.

Moffat A.M., Papale D., Reichstein M., Hollinger D.Y., Richardson A.D., Barr A.G., et al., (2007) Comprehensive comparison of gap filling techniques for eddy covariance net carbon fluxes, *Agricultural and Forest Meteorology*, 147, 209-232.

Muehrcke P.C., Muehrcke J.O., (1992) Map use: reading, analysis, Interpretation. Madison W.I., J.P. Publications.

Murtagh F., (1995) Interpreting the Kohonen self-organizing map using contiguity-constrained clustering, *Pattern Recognition Letters*, 16, 399-408.

Vesanto J., Alhoniemi E., (2000) Clustering of the self-organizing map, *IEEE Transactions on neural networks*, 11(2), 586-600.

Walter H., Box E., (1976) Global Classification of Natural Terrestrial Ecosystems, *Vegetatio*, 32, 75-81.

Zang X., Li Y., (1993) Self-organizing map as a new method for clustering and data analysis. In *Proceedings of international joint conference on neural networks* (IJCNN'93), 2448-2451.

Ringraziamenti

Quando si decide di intraprendere un corso di dottorato 10 anni dopo una laurea e alla soglia degli "anta", è inevitabile che alla fine del lavoro siano molte le persone che ci si sente in dovere di ringraziare:

il primo ringraziamento va a Dario, per il suo supporto e soprattutto per la pazienza con cui ha saputo "rimettere in moto" la macchina e spingerla ogni qual volta essa tentava di fermarsi;

grazie a Riccardo, per avermi dato la possibilità di iniziare e concludere questo cammino;

grazie a Sebastiaan Luyssaert e a Mika Sulkava, per avermi aperto il mondo delle SOM e per avermi accompagnato nella loro applicazione con preziosi suggerimenti e correzioni;

grazie a Monia, instancabile appassionata di GIS, che con la sua competenza e professionalità mi ha introdotto nel mondo di ArcGis e che ha avuto la pazienza di rispolverare con me tutte le competenze ormai assopite;

grazie a Carlo, per il suo tempo perso dietro le mie innumerevoli richieste e per il supporto dato con i codici di Matlab con cui sono stati preparati i dati spazializzati e quelli dei siti eddy;

grazie a tutti quelli che, anche senza accorgersene, mi hanno sostenuto in questo ultimo periodo anche solo ascoltandomi e dedicandomi un po' di attenzione e tempo;

grazie a mia madre per tutti i pranzi, le cene e il resto delle cose che ha fatto al mio posto!!

E infine, ma non certo per ultimi, un profondo grazie e un bacio vanno a Marcello e Michele, che hanno sopportato pazientemente le mie giornate, le mie serate e le mie assenze, per i quali non servono parole… loro sanno…